FOIA Processing in the Presidential Electronic Records PilOt System

PERPOS Working Paper ITTL/CSITD 06-05
July 2006

Sandra Laib
William Underwood

Computer Science and Information Technology Division
Information Technology and Telecommunications Laboratory
Georgia Tech Research Institute
Georgia Institute of Technology

The Army Research Laboratory (ARL) and the National Archives and Records Administration (NARA) sponsor this research under Army Research Office Cooperative Agreement DAAD19-03-2-0018. The findings in this paper should not be construed as an official ARL or NARA position unless so indicated by other authorized documentation

ABSTRACT

The Presidential Electronic Records Pilot System (PERPOS) supports systematic processing of Presidential Electronic Records. It consists of two software components. The Archival Repository Tool (ART) supports accession, cataloging, repository management, and production of reference copies of e-records that are available for public access. The Archival Processing Tool (APT) supports arrangement, preservation, and review of electronic records in the repository.

This report describes additional functions developed for PERPOS that support FOIA Processing of Presidential e-records. FOIA Processing is initiated by a citizen's request for records under the Freedom of Information Act (FOIA). A search is made of accessioned records (usually unprocessed) to determine the records related to the request. The requestor is notified of the volume of records (in pages) that are potentially relevant and an estimate is made of the time needed to process them. An archivist will then review just those records that might be relevant, not considering a series in it's entirety, but just relevant records in several series.

The functionality added to PERPOS that supports FOIA processing includes:

- Indexing the accessioned electronic records,
- Creating a FOIA case,
- Searching the indexed records for records relevant to a FOIA Request,
- Automatic estimation of the number of pages of e-records relevant to a request,
- Reviewing records for a FOIA case,
- Creating the Scope and Content Note for a FOIA case, and
- Automatically creating a finding aid and a container for a FOIA Collection.

Keywords: FOIA, Presidential E-records

Table of Contents

1. INTRODUCTION 1

1.1 BACKGROUND 1
1.2 PURPOSE 1
1.3 SCOPE 1

2. FOIA PROCESSING DATAFLOW 2

3. FOIA REQUESTS, INDEXING AND INITIAL SEARCHES 3

3.1 INDEXING CONTAINERS 4
3.2 FOIA CASE MANAGEMENT 5
3.3 PERFORMING A SEARCH 7

4. ESTIMATING THE NUMBER OF PAGES TO BE REVIEWED 12

5. REVIEWING RECORDS FOR FOIA CASES 14

5.1 CHECKOUT A CONTAINER 14
5.2 OPEN A CONTAINER FOR REVIEW 16
5.3 REVIEW RECORDS RELATED TO FOIA CASE 18
5.4 ADD TO OR REMOVE RECORD FROM A FOIA CASE 23
5.4.1 Add a Record to a FOIA Case 23
5.4.2 Remove a Record from a FOIA Case 24
5.5 CHECK IN A CONTAINER 25
5.6 UNDO CHECKOUT 27

6. FOIA CASE DESCRIPTION 29

7. FOIA REFERENCE CONTAINER, ITS ARRANGEMENT AND FINDING AID 29

7.1 FOIA REFERENCE CONTAINER AND FINDING AID ARRANGEMENT 30
7.2 MAKE A FOIA REFERENCE CONTAINER AND FINDING AID 30
7.3 OPEN A FOIA REFERENCE CONTAINER 31
7.4 OPEN A FOIA COLLECTION FINDING AID 33

8. INACTIVE FOIA CASE FILES 34

9. CHANGES TO SYSTEMATIC PROCESSING 35

9.1 CREATING A SYSTEMATIC CASE 35
9.2 ADD SYSTEMATIC CASE CONTAINER 36
9.3 SYSTEMATIC PROCESSING OF CONTAINERS 37

10. SUMMARY AND RELATED RESEARCH 37

REFERENCES 41

APPENDIX A: LIBRARY REFERENCE REQUEST FORM 42

APPENDIX B: LIBRARY REFERENCE SEARCH FORM 43

APPENDIX C: SAMPLE FINDING AID FOR A FOIA CASE 44

APPENDIX D: CHANGES TO THE REPOSITORY DATABASE TABLES 46

APPENDIX E: CHANGES TO ART AND THE APT TO ENABLE FOIA PROCESSING 51

1. Introduction

1.1 Background

The Archival Processing Tool (APT) and Archival Repository Tool (ART) prototypes developed in prior research support systematic processing of Presidential electronic records. An archivist systematically processes an accessioned record series by considering the arrangement, preservation, review, and description of the entire series. In the early years of operation of a Presidential Library, archivists primarily perform FOIA processing, rather than systematic processing. FOIA processing is initiated by a citizen's request for records under the Freedom of Information Act (FOIA). A search is made of accessioned records (usually unprocessed) to determine which records are relevant to the request. The requestor is notified of the volume of records (in pages) that are relevant and an estimate is made of the time needed to process them. An archivist will then review just those records that might be relevant, usually not considering a series in it's entirety, but just folders in several series that contain relevant records. While an archivist may also perfect the arrangement and perform preservation actions on those records reviewed, they usually do not fully describe, preserve or arrange entire series.

1.2 Purpose

The purpose of this report is to describe

- the functions developed for PERPOS that support FOIA processing of Presidential electronic records,
- the Systematic Case Processing feature that was added to distinguish FOIA Processing from Systematic Processing, and
- the database schema that was added to the Repository to support FOIA Processing.

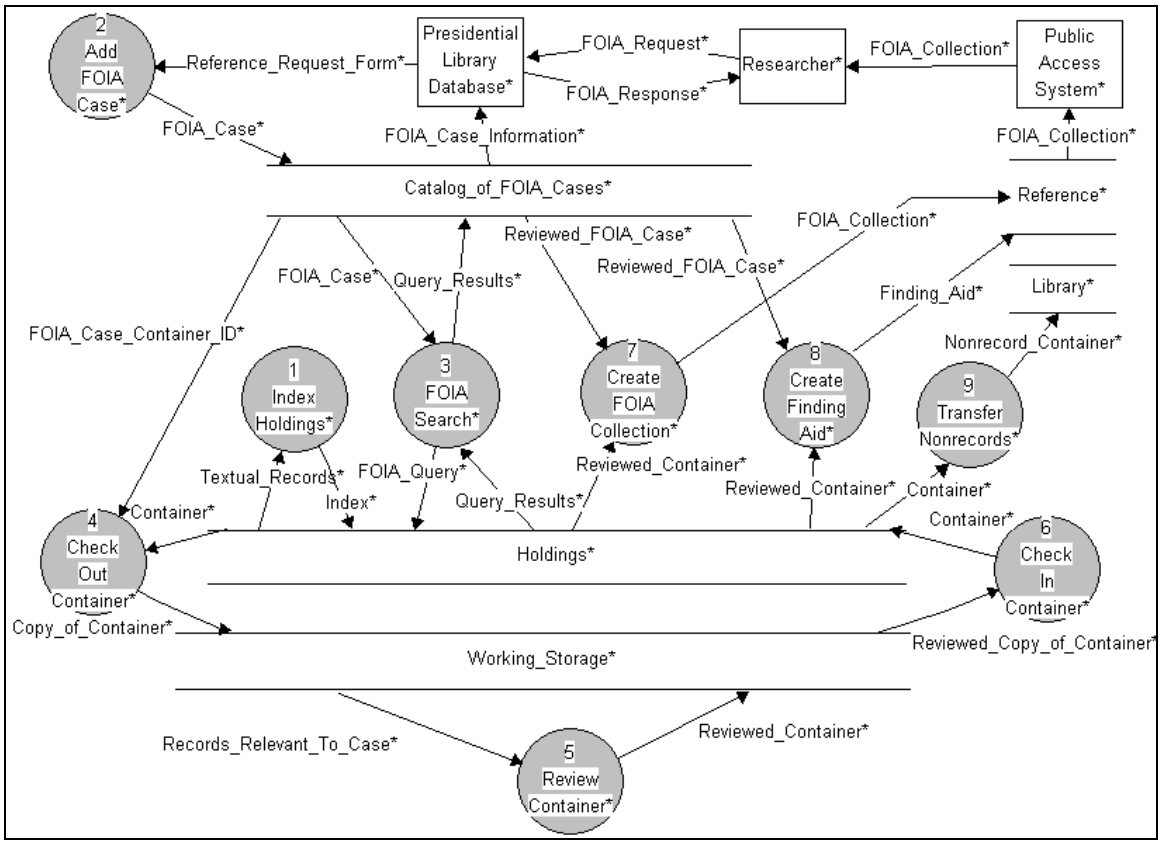
1.3 Scope

The next section describes the dataflow of FOIA processing. The third section describes the functions of indexing accessioned electronic records, creating a FOIA Case, and searching the indexed records for records relevant to a FOIA Request. The fourth section describes the method used for estimating the number of pages of electronic records relevant to a request. The fifth section describes the process of reviewing records for a FOIA case. Section 6 describes the process of creating the Scope and Content Note for a FOIA Collection. Section 7 describes the process of automatically creating a finding aid and a container for a FOIA Collection. Section 8 discusses alternatives for disposition of inactive FOIA Collections. Section 9 describes the addition to PERPOS of a Systematic Case Management function. Section 10 describes planned enhancements. Appendices A-C show sample forms used by the Bush Library for Processing FOIA Requests and a

sample finding aid for a FOIA Collection. Appendices D and E describe database schema that were added to the Archival Repository Tool (ART) to support FOIA Processing and the addition of manifest library and page estimation components to ART.

2. FOIA Processing Dataflow

The diagram below illustrates the dataflow of FOIA processing activities that are supported by the PERPOS tools. The numbered, labeled circles are activities supported by the tools. The two parallel lines represent data stores. The rectangles represent entities external to the PERPOS system, and the labeled, directed arrows are data flows. Stepping through the diagram in the numerical sequence of the activities, one sees the dataflow.



Holdings consist of accessioned containers of e-records. The PERPOS Tools first support FOIA processing by providing the capability to Index Holdings (Activity 1). This function creates an index of all the terms in the textual records in Holdings.

The Bush Presidential Library Database is a Microsoft Access database of tables, forms, reports, and queries. The database includes an accession register, location register, folder title list, Reference Request Form and a Reference Search Form. The Presidential Library Database is external to the PERPOS Tools. When a request for records under the Freedom of Information Act is received from a Researcher, a Reference Request Form is filled out and a FOIA Case number assigned. The various collections are searched for

records relevant to the Request and the results of these searches are entered into a Reference Search Form. To perform a search of the Staff Member and Office electronic records in Holdings, a FOIA case is created (Activity 2) that has the same FOIA Case Number as that on the Reference Request Form. An archivist translates the FOIA request into a FOIA Query, which is submitted to FOIA Search (Activity 3) that returns a result set of pointers to records that are relevant to the FOIA query. The result set is associated with the FOIA case. The requestor is notified of the volume of records (in pages) that is relevant and an estimate is made of the time needed to process them (not shown in the data flow). When an archivist is ready to begin work on the FOIA case, they check out one of the containers associated with the case (Activity 4) to an archivist's work area. The archivist then uses the review activity of the Archival Processing Tool (APT) to review just those records in the container that are relevant to the query, not considering an entire record series or container (Activity 5). Review actions include opening a record for public access; withdrawing or redacting a record because of access restrictions; marking a record as a Personal Record Misfile or marking a file for transfer to the Library because it is a non-record. When the records that are in the result set for a container are reviewed, the archivist checks the container back into Holdings (Activity 6). Once all containers with relevant e-records are reviewed, the Archival Repository Tool is used to make a FOIA Reference Collection (Activity 7) and Finding Aid (Activity 8). Records Marked for Transfer are transferred to the Library (Activity 9). The researcher is notified of the completion of the review, the availability of the requested records, and may be sent a copy of the finding aid. The Public Access System has not yet been developed.

The following sections describe each of these support functions in detail and show the user interface.

3. FOIA Requests, Indexing and Initial Searches

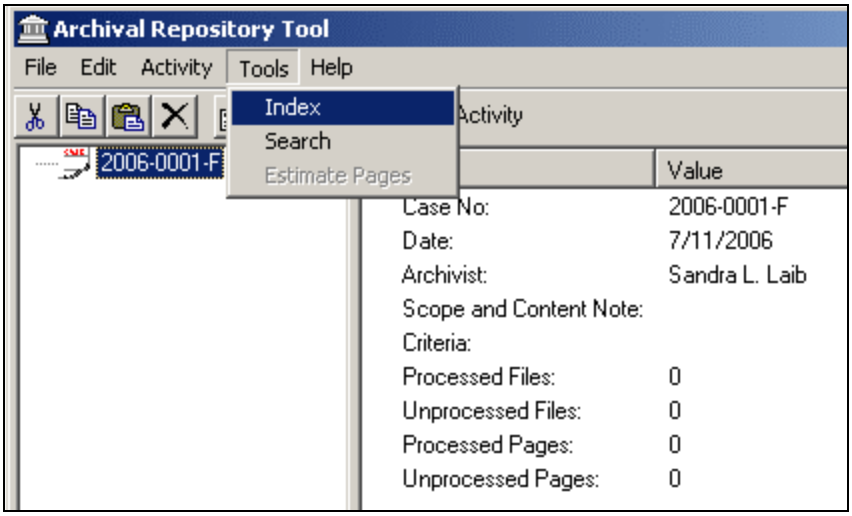
FOIA requests are logged into the Presidential Library Database using a Bush Presidential Library Reference Request Form (See Appendix A). A paper copy of this form goes into a yellow folder labeled with the requestor’s last name and the assigned FOIA case number. Then all applicable databases are searched for records relevant to the request. The results are logged on a Bush Presidential Library Reference Search Form (See Appendix B) and a paper copy of this form is placed in the same yellow folder.

The Bush Presidential Library searches the Staff Member and Office Files at the folder level using a folder title list. The Bush Presidential Library archivists process Staff Member and Office Files (of which the Bush PC files are examples) at the folder level. Even if some of the records in a folder are not relevant, the entire folder is processed. This does not apply to the WHORM Subject or Alpha Files where records are tracked on an individual basis and processed individually.

3.1 Indexing Containers

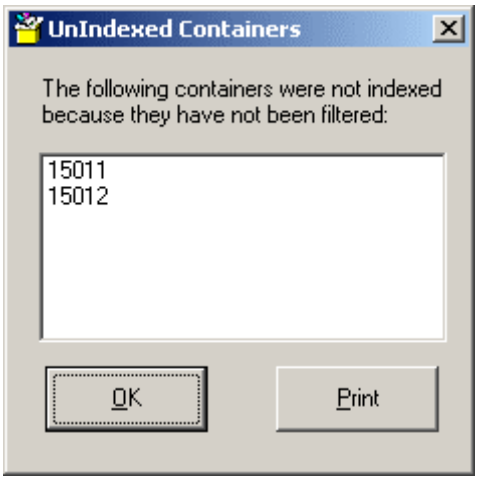
The Archival Repository Tool is used to index the accessioned containers of electronic records that are in the Repository (Holdings). Only the records in containers that have been filtered, or that have been marked as filtered, are indexed. Filtering is the process of removing operating system files, office software application files, system or software documentation, or sample application files from accessioned records. These files are not Presidential records and indexing them would result in an archivist having to consider non-records during the review process. Filtering of containers of record series is also necessary for Systematic Processing and is discussed in the PERPOS Tools User Manual [Underwood et al 2006].

To index the contents of all accessioned and filtered containers, the *FOIA Case Mgmnt* Activity must be the current activity. To make this the current activity, the archivist selects *FOIA Case Mgmnt* from the *Activity* drop-down menu. Next, the archivist selects *Index* from the *Tools* drop-down menu.



The message "Indexing Containers …" will appear in the status bar at the bottom of the screen. Depending on the number of containers and number of files in each container, this process will take from a few minutes up to several hours. If any containers have not been filtered, a dialog box will appear that lists the OAID[1] of the containers that were not indexed because they were not marked as filtered.

[1] The abbreviation OAID stands for "Over-Sized Attachment ID." This is a 5-digit number assigned by the White House Office of Records Management (WHORM) to the Federal Record Center (FRC) boxes of records transferred to the National Archives. The Presidential Library assigns additional OAID numbers to other accessions. The archivists originally suggested that 5-digit OAID numbers be assigned to the containers of electronic records accessioned by the Library. They now suggest some other name for identifiers of containers of digital materials, e.g., Digital Container ID, to distinguish them from FRC boxes.



The archivist can get a printout of these container ids by selecting the *Print* button at the bottom of the dialog box. Select the *OK* button at the bottom of the form to complete the process.

It is necessary to re-index the containers in the holdings area in the following cases.

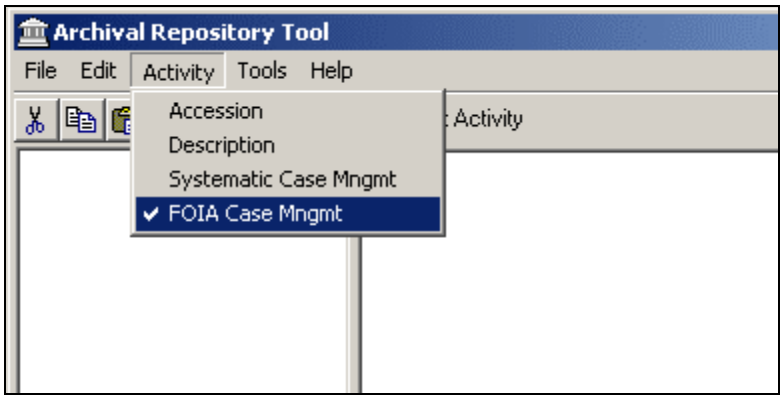
(1) Since the index was last created, new containers have been accessioned and/or filtered.
(2) Since the index was last created, non-record files have been transferred from containers to the Library during the arrangement, preservation, or review process.
(3) Removal of Personal Record Misfiles (PRMs).

In the first case, records relevant to a FOIA search may have been added to Holdings, but will not appear in a result set because they are not indexed. In the second case, a FOIA search may return a pointer to a record that is no longer in its original container. In the third case, a FOIA search may return a pointer to a record that is no longer in its original container because upon review it is found to be a PRM and is subsequently transferred out of the container. It is not necessary to re-index holdings before each search of holdings, if these conditions have not occurred.[2]

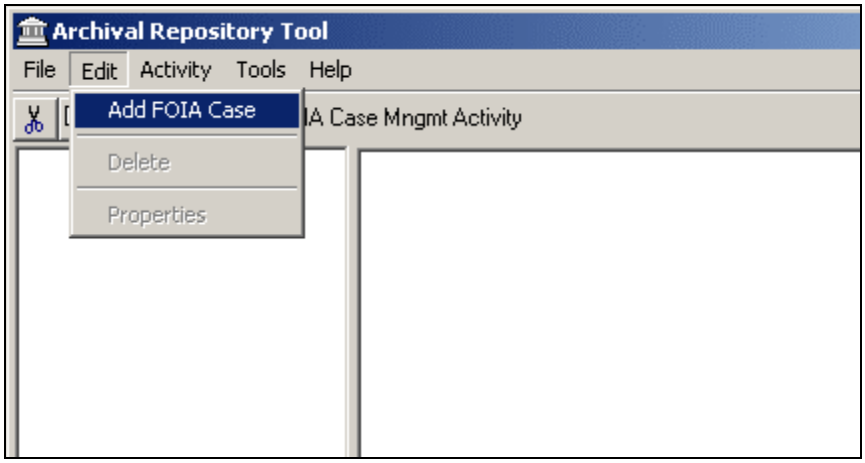
3.2 FOIA Case Management

The Archival Repository Tool is used to manage the search for and processing of electronic records relevant to a FOIA Case. To do this, select *FOIA Case Mgmnt* from the *Activity* drop down menu shown below.

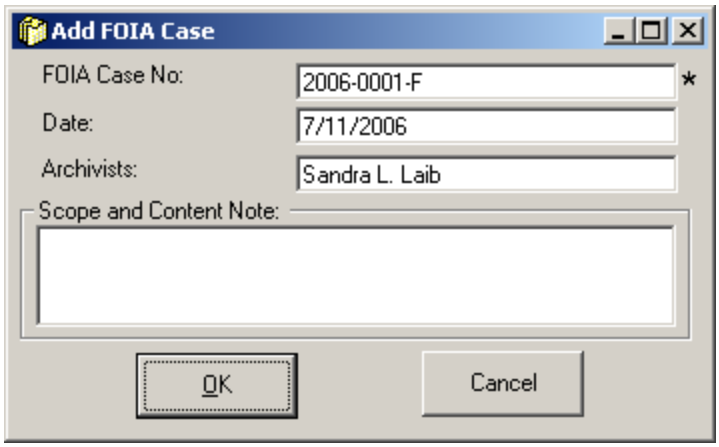
[2] When archivists attempt a FOIA search, they are notified when then need to reindex holdings. A container table has a field that indicates the date a container was last indexed (or is unindexed), and the date when files were last transferred out of the container, including the date of removal of PRMs.



The archivist then selects *Add FOIA Case* from the *Edit* drop down menu.



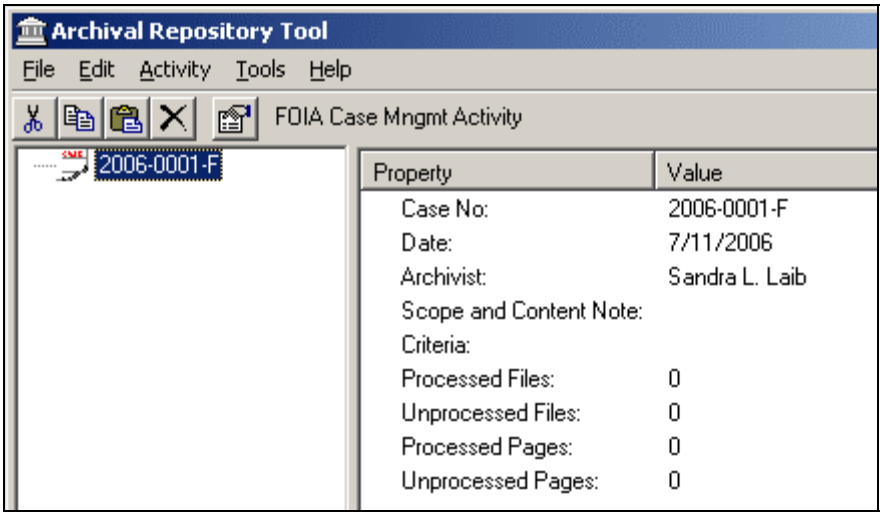
This causes the Add FOIA Case dialog box to appear.



The Archival Repository Tool automatically fills in the Archivist field with the user name and the *Date* field with the system date. In the dialog box, the archivist enters the FOIA Case number from the Bush Presidential Library Reference Search Form and selects *OK*. There is a check to be sure that there is not a collision with an existing FOIA case

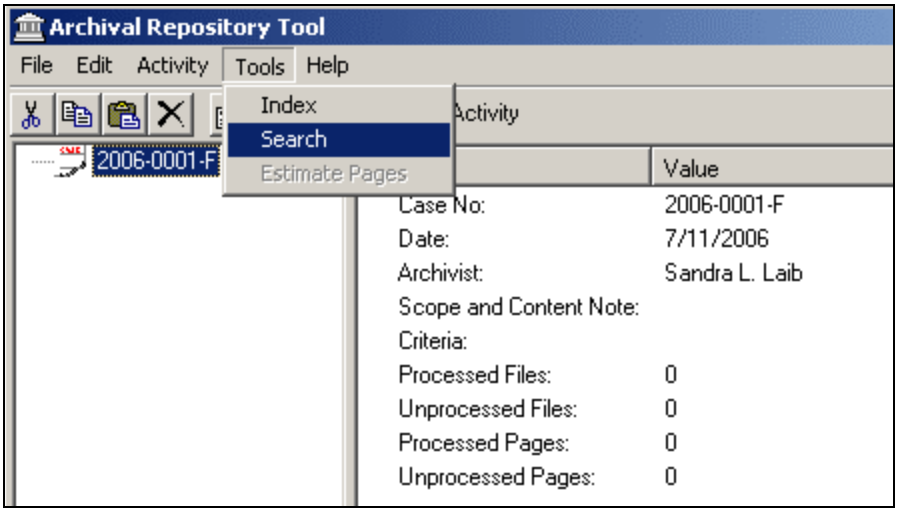
number. The *Scope and Content Note* should not be filled out until the FOIA Case is complete and it is time to create a FOIA Reference copy.[3]

After *OK* is selected, the archivist is returned to the *FOIA Case Mngmt* activity with the FOIA Case Number highlighted in the left pane and the FOIA Case properties in the right pane.



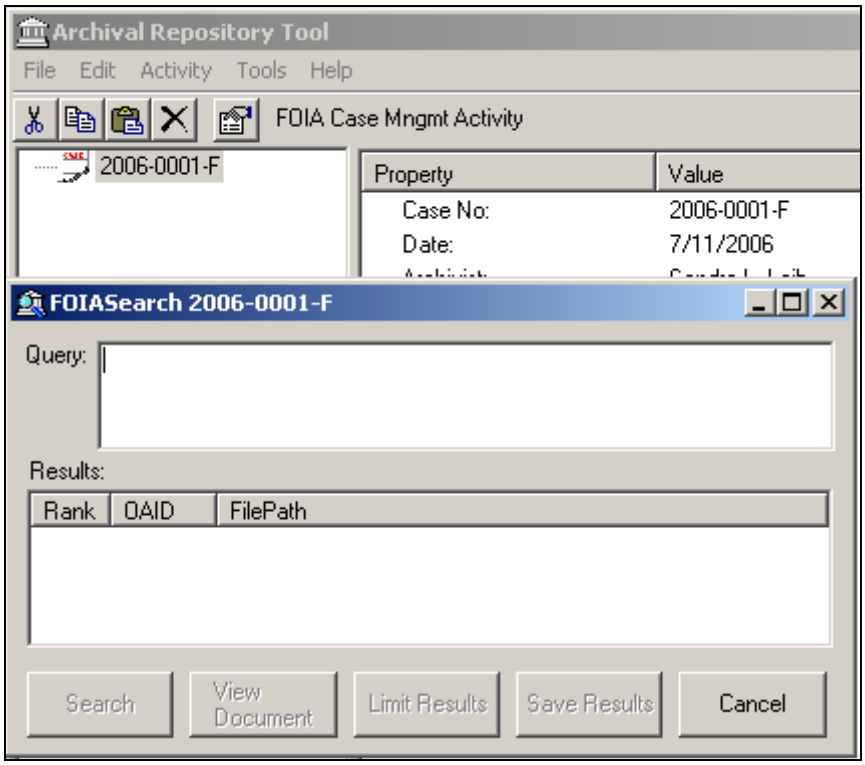
3.3 Performing a Search

To search for records relevant to a FOIA request, the *FOIA Case Mngmt* Activity must be the current activity. To begin a FOIA Search, the archivist must first select a FOIA Case. When a FOIA Case is highlighted, the archivist selects *Search* from the *Tools* drop-down menu.

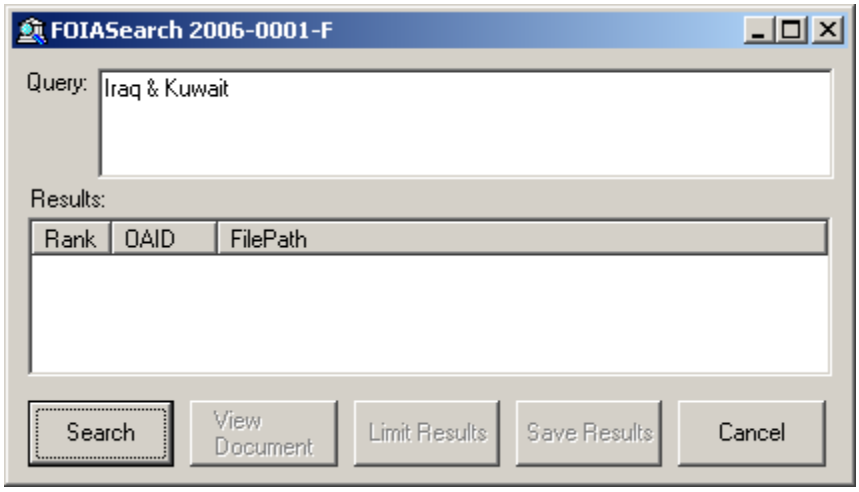


[3] A check could be made as to whether review of the FOIA case was complete. If not, it would not be possible to enter the Scope and Content Note.

This starts the FOIASearch application. The FOIASearch's main screen appears with the FOIA Case No in the screen title.



The archivist translates the FOIA request from the Reference Request Form into an Oracle Text with word query. For instance, if the request had been for "Any materials related to Iraq and Kuwait," the Oracle Text query might be "Iraq & Kuwait." The archivist enters the search criteria into the Query box.



Oracle Text uses the Boolean operators AND (&), OR (|) and NOT (~). Parentheses are used for grouping expressions. For example '(George | Barbara) & Bush' is for all records containing the words either 'Barbara' or 'George' and the word 'Bush'.

A root word prefixed with a dollar sign ($), e.g., $broadcast, will find all documents containing its root word (stem) or derivatives, e.g., broadcasts, broadcasting, or broadcaster. The EQUIV operator (=) is used to indicate that two or more words are equivalent, for instance (91=1991).

Using the ACCUM(ulate) (,) and weight (*) operators, one can increase the score for documents that match a query by weighting terms differently. For instance, in searching for documents related to the *Clarence Thomas nomination to the Supreme Court*, the expression

(justice, judge, Supreme Court*5, Clarence Thomas *10)

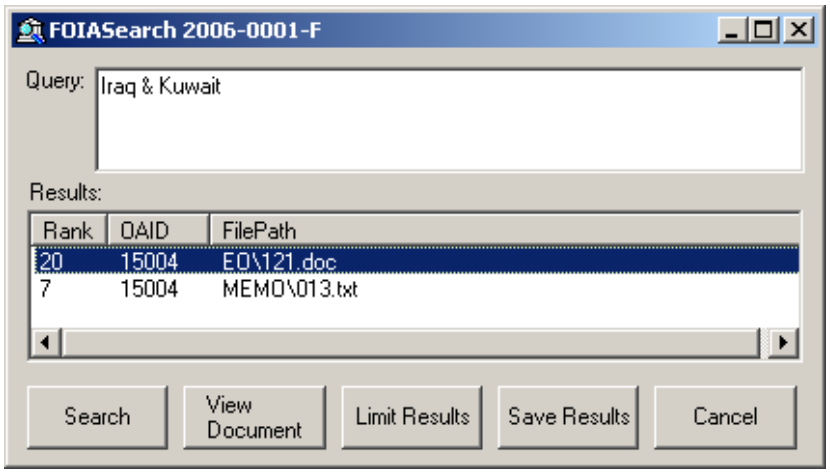
will increase the score of the term *Supreme Court* by 5 times and the term *Clarence Thomas* by 10 times. This signifies that documents related to *Clarence Thomas* and *Supreme Court* are most relevant to the query. The ACCUM operator gives the highest scores to documents that contain the terms within the scope of the operator; for instance, ACCUM (dog, pet, Millie) will give the highest score to documents that contain all three terms.

One can search for terms that are in close proximity with the NEAR operator. For example, to find all documents where Soviet is within six words of Revolution, the following query would be issued.

NEAR((Soviet, revolution), 6)

The default and maximum value for the NEAR operator is to search for terms separated by no more than 100 words. In conjunction with Boolean operators, the NEAR operator constrains the scope of a query. Used with the section-searching operator WITHIN, the NEAR operator can constrain the search to predefined zones (sentence, paragraph, HTML sections).

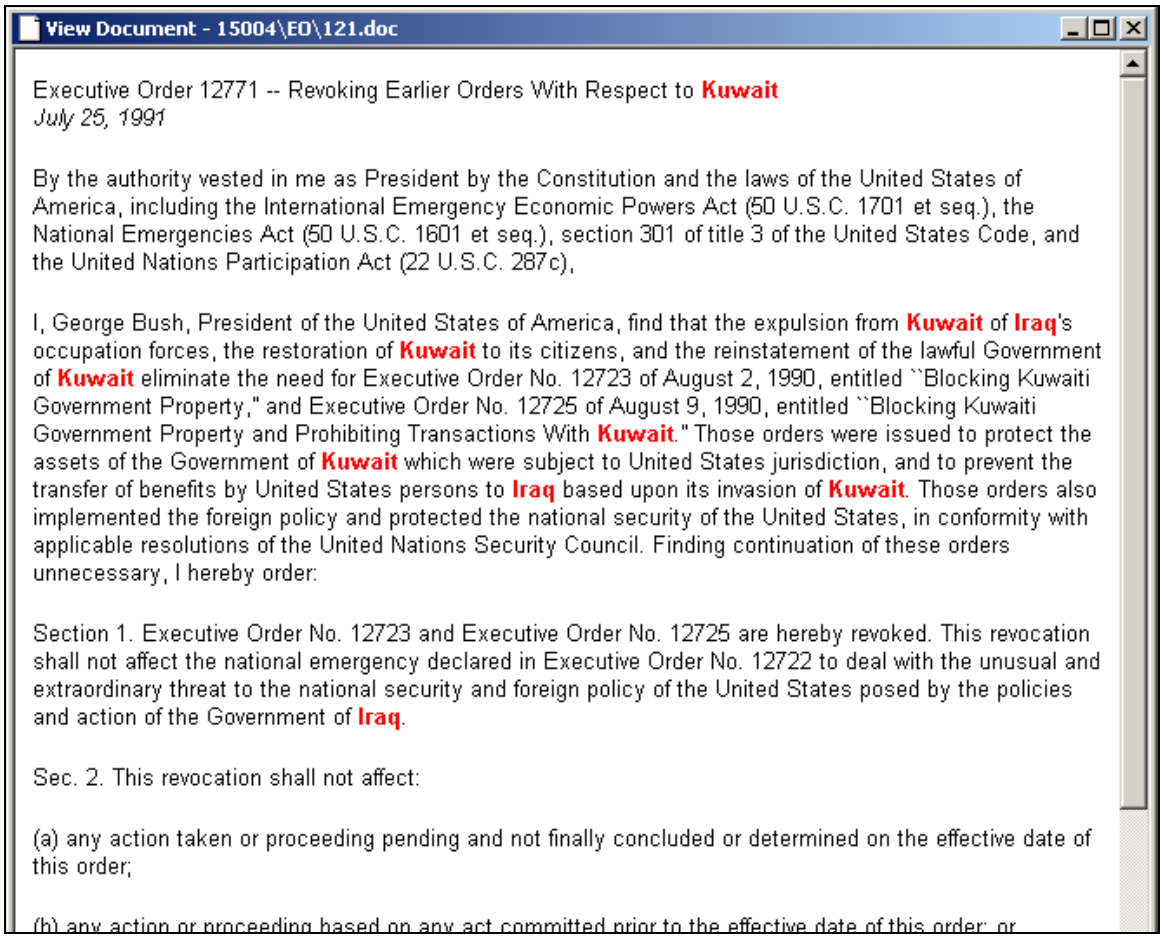
Once the query is entered, the archivist selects the *Search* button.



Once a search has been performed, the *View Document*, *Limit Results*, and *Save Results* buttons are enabled. The result set is displayed as a table with the column headings Rank, OAID, and FilePath. The Rank is the relevance value assigned by Oracle Text Search. To calculate the relevance score for a record that matches a query, Oracle Text uses an inverse frequency algorithm based on Salton's formula.[4] Inverse frequency scoring assumes that for a record to score high, the query term must occur frequently in the record, but infrequently in the entire record set.

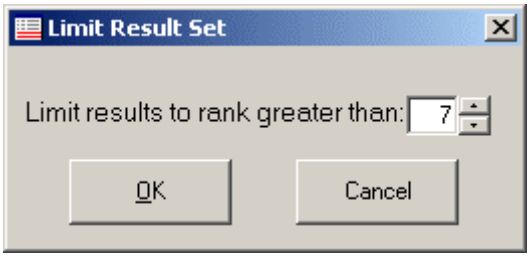
The OAID is the unique identifier the archivist assigned to the container when it was accessioned. The FilePath is the path within the container to the file (record).

To determine whether a record is relevant to the FOIA request, an archivist performing the FOIA search is able to view individual records with the search terms highlighted. The View Document screen also displays the OAID along with the path within the container in the title of the screen. To do this, the archivist selects the row in the table containing the path to the file he/she wishes to view and selects the *View Document* button.


View Document - 15004\EO\121.doc
Executive Order 12771 -- Revoking Earlier Orders With Respect to Kuwait
July 25, 1991
By the authority vested in me as President by the Constitution and the laws of the United States of America, including the International Emergency Economic Powers Act (50 U.S.C. 1701 et seq.), the National Emergencies Act (50 U.S.C. 1601 et seq.), section 301 of title 3 of the United States Code, and the United Nations Participation Act (22 U.S.C. 287c),
I, George Bush, President of the United States of America, find that the expulsion from Kuwait of Iraq's occupation forces, the restoration of Kuwait to its citizens, and the reinstatement of the lawful Government of Kuwait eliminate the need for Executive Order No. 12723 of August 2, 1990, entitled ``Blocking Kuwaiti Government Property,'' and Executive Order No. 12725 of August 9, 1990, entitled ``Blocking Kuwaiti Government Property and Prohibiting Transactions With Kuwait.'' Those orders were issued to protect the assets of the Government of Kuwait which were subject to United States jurisdiction, and to prevent the transfer of benefits by United States persons to Iraq based upon its invasion of Kuwait. Those orders also implemented the foreign policy and protected the national security of the United States, in conformity with applicable resolutions of the United Nations Security Council. Finding continuation of these orders unnecessary, I hereby order:
Section 1. Executive Order No. 12723 and Executive Order No. 12725 are hereby revoked. This revocation shall not affect the national emergency declared in Executive Order No. 12722 to deal with the unusual and extraordinary threat to the national security and foreign policy of the United States posed by the policies and action of the Government of Iraq.
Sec. 2. This revocation shall not affect:
(a) any action taken or proceeding pending and not finally concluded or determined on the effective date of this order;
(b) any action or proceeding based on any act committed prior to the effective date of this order; or

[4] $3f(1+\log(N/n)$ where f is the frequency of the query term in the document, N is the number of documents, and n is the number of documents containing the query term.

The archivist can limit the results of the query to any result with a rank greater than one of their choosing. To do this, the archivist selects the *Limit* Results button.



However, in this case, it is not necessary to limit the result since the second record is relevant. The archivist can remove all results that do not have rank greater the seven by selecting the *OK* button. The archivist can also choose to change the search criteria by simply changing the query and selecting the search button. The old results will be discarded and the results of the new search will be displayed.

When the archivist selects the *Save Results* button at the bottom of the Search screen, the query is returned to the Archival Repository Tool along with the query results and is associated with the appropriate FOIA Case for further processing.[5] At any time, the archivist can leave the FOIASearch application by selecting the *Cancel* button at the bottom of the dialog box. If this button is selected, neither the query or result list will be returned to the Archival Repository Tool and associated with the FOIA case.

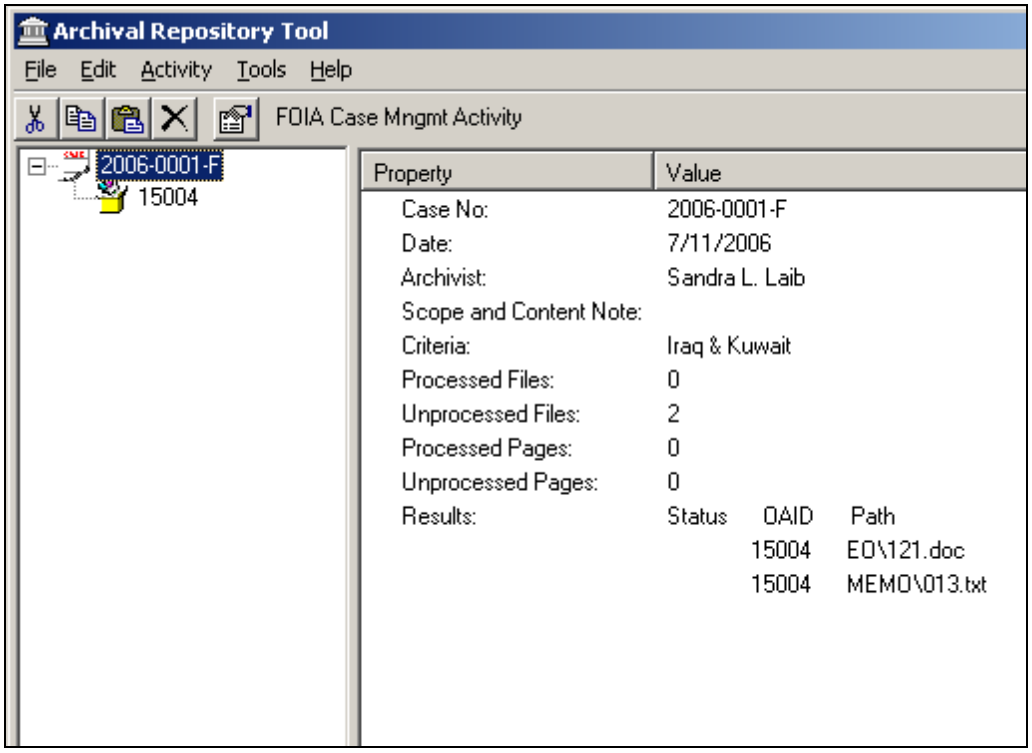
If the FOIA results search is saved, both the query and the results will be displayed in the right windowpane of the Archival Repository Tool when the FOIA Case is highlighted. The containers that contain the relevant records will be listed under the FOIA Case. The results will be displayed in a list with the headings 'Status', 'OAID', and 'Path'.[6] The status of a record can be 's', 'f', or blank. The status will be 's' for a record that has been systematically processed, so will not need to be reviewed. It will be 'f' for a record that has been processed for a previous FOIA case, so will not need to be reviewed. It will be blank for a record that is unprocessed, so will need to be reviewed.[7] Under the result list,

[5] Currently, the FOIA Case Management activity and FOIA Search capability only support saving the result set of one query for a FOIA Case. For complex FOIA requests or for follow-up searches, it will be necessary to have the capability for additional queries and result sets to be associated with a FOIA case. It is planned to accomplish this by extending the FOIA Case No to have subcases, e.g., 2006-0020-F[1], 2006-0020-F[2], and creating FOIA search criteria and results sets for each subcase. (See Appendix D, FOIACase Table)

[6] It is not clear that this information (Status, OAID, and Path) needs to be displayed. It is needed in the results table by ART and the APT to identify the containers that have records that need to be reviewed, the specific records that need to be reviewed, and their status for preparation of the finding aid. An archivist processing paper records in folders needs such a list because they have to manually find particular folders in particular containers. However, that manual capability is not needed in this FOIA processing environment.

[7] This status information is used to calculate the number of processed and unprocessed files, which gives an archivist an idea of the amount of work needed to fulfill the FOIA request. This information is also used, after review of the unprocessed records, to prepare the Finding Aid for the FOIA collection.

the properties *Processed Files*, *Unprocessed Files*, *Processed Pages*, and *Unprocessed Pages* are listed. The Archival Repository Tool calculates the *Processed Files* and *Unprocessed Files* properties. These property values are updated each time a container is returned to holdings, even if the container was checked out for another FOIA case or systematic processing.



4. Estimating the Number of Pages to be Reviewed

An archivist responding to a FOIA request must enter into the Reference Search Form the estimated number of pages to be processed. Files of the same size but different file format can correspond to very different numbers of pages. For example, a Microsoft Excel file may take up a relatively small amount of file space, but generally converts to a large number of pages. Whereas, an image file may have a large number of bytes but correspond to a single page. Another factor is the size of the header that some file formats have.

For the file types that occur in the Bush personal computer files, our approach is to estimate the average number of bytes per page for each file format type t.

$$\text{Avgbytesperpage}(t) = \left(\sum_{i=1,n} (\text{filesize}(i) - \text{headerlength}(i))/\text{number of pages}(i)\right)/n$$

where n is the sample size of the records of type t.

The average number of bytes per page for each file type along with any standard header length is stored in a table called PageEstimates. Given some files that have been determined to be relevant to a FOIA request, for each file i of type t

$$Estpages(i, t) = (filesize(i) - headerlength(t))/Avgbytesperpage(t)$$

For some file types, such as charts, where it is known that there is only one page per file, no calculation is necessary (Estpages(i,t) = 1).

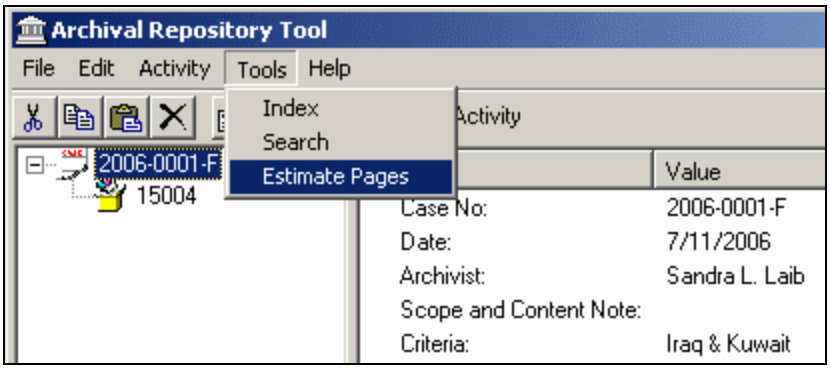
Accumulate the estimated number of pages that are processed (or unprocessed) for a FOIA case as

$$Estpagescase = \sum_{i=1,n} Estpages(i, filetype(i))$$

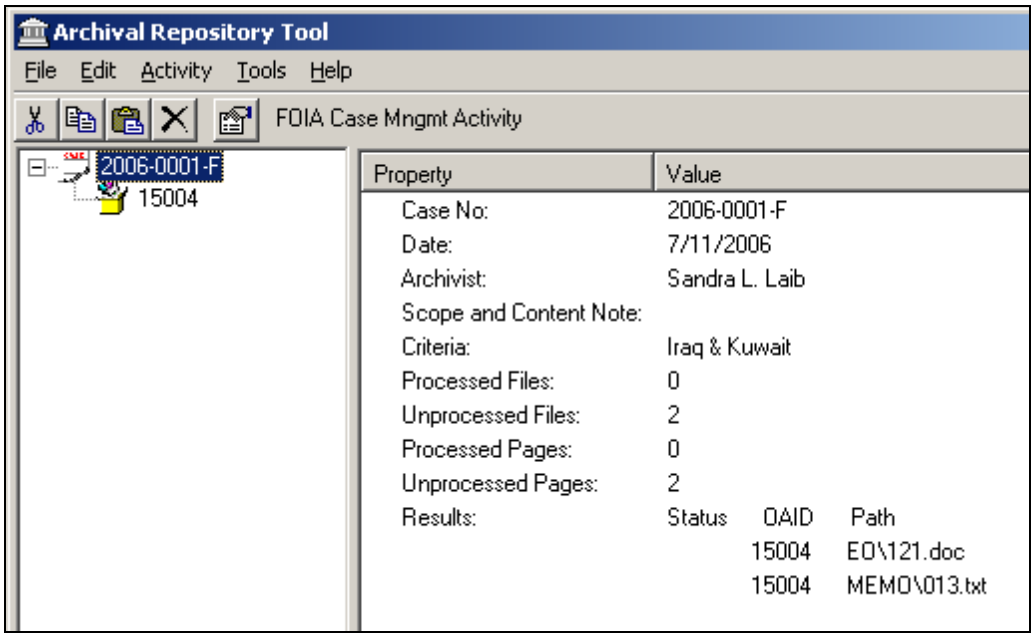
where n is the number of files in a case that are processed (or unprocessed).

The results are stored as properties of the FOIA Case and displayed in the right pane of a selected FOIA Case.

To have the Archival Repository Tool estimate the number of pages for a FOIA Case, the *FOIA Case Mngmt* Activity must be the current activity. To calculate the estimated number of processed and unprocessed pages, the archivist must first select a FOIA Case. When a FOIA Case is highlighted, the archivist selects *Estimate Pages* from the *Tools* drop-down menu.



When the Archival Repository Tool has estimated the number of pages in the FOIA Case, it updates the currently selected FOIA Case and displays the new estimates in the right pane.



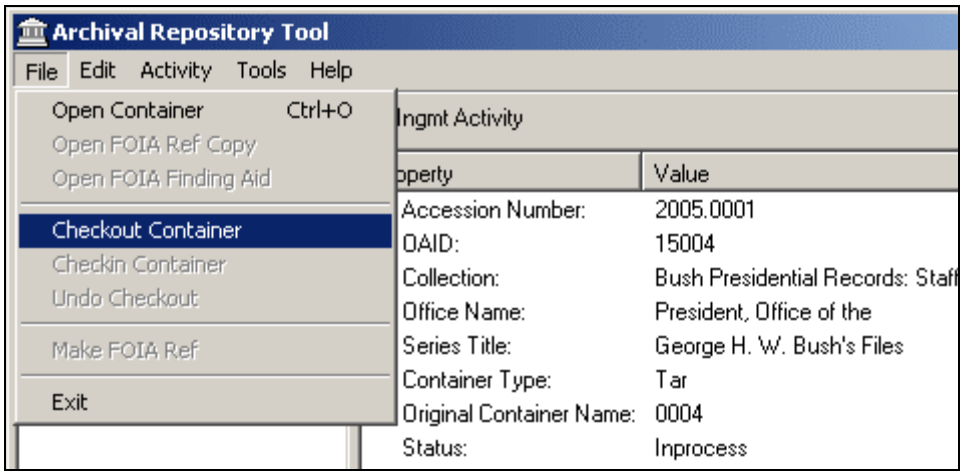
5. Reviewing Records for FOIA Cases

To begin the review of records relevant to a FOIA case, an archivist opens the Archival Repository Tool. The archivist then selects *FOIA Case Mngmt* from the *Activity* drop-down menu. The archivist selects the FOIA case to be processed. The FOIA case's properties appear in the right panel. If FOIA search is performed, the query criteria and the result list will appear in the property list. Lists of containers in which there are records relevant to the request are displayed beneath the FOIA Case in the left pane. Click on the '+' symbol beside the FOIA case, to view the list of containers. The archivist then selects a container to review. If the container is already checked out for processing by another archivist, the *Archivist, Processing Type,* and *Case No* properties will appear in the Container's property list in the right panel. Containers can only be checked out to one archivist at a time.

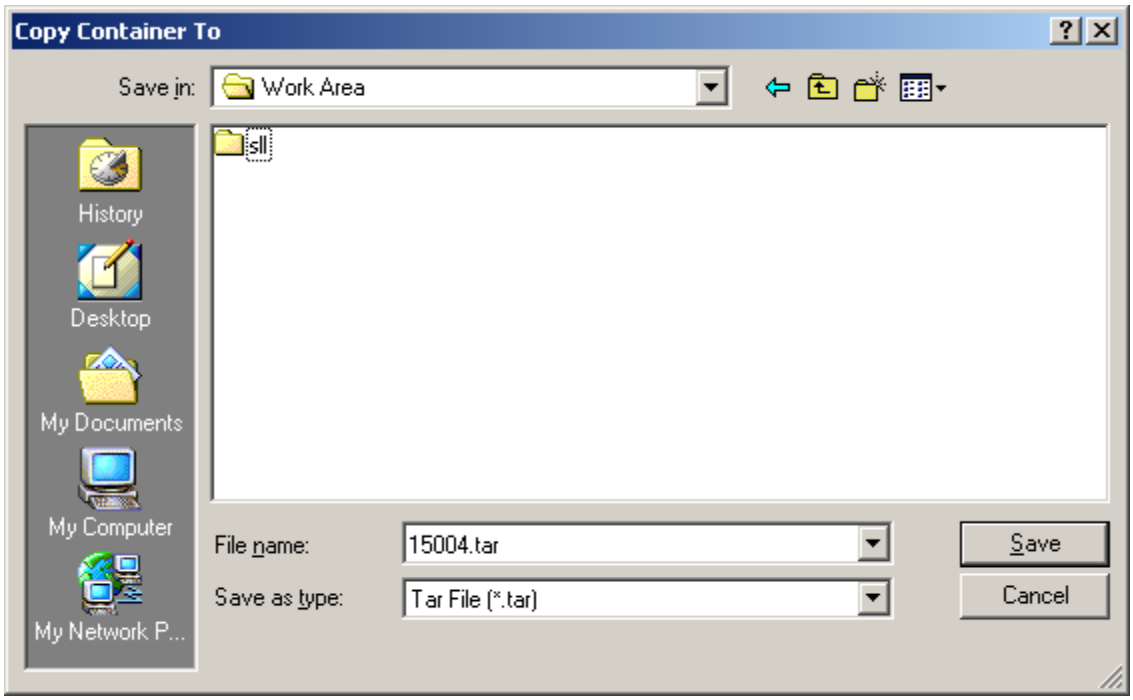
5.1 Checkout a Container

If the container is not already checked out to another archivist, the archivist checks it out for processing by selecting *Checkout Container* from the *File* drop-down menu. There is a check to determine that the archivist who is checking out the container is the archivist who created the FOIA case.[8]

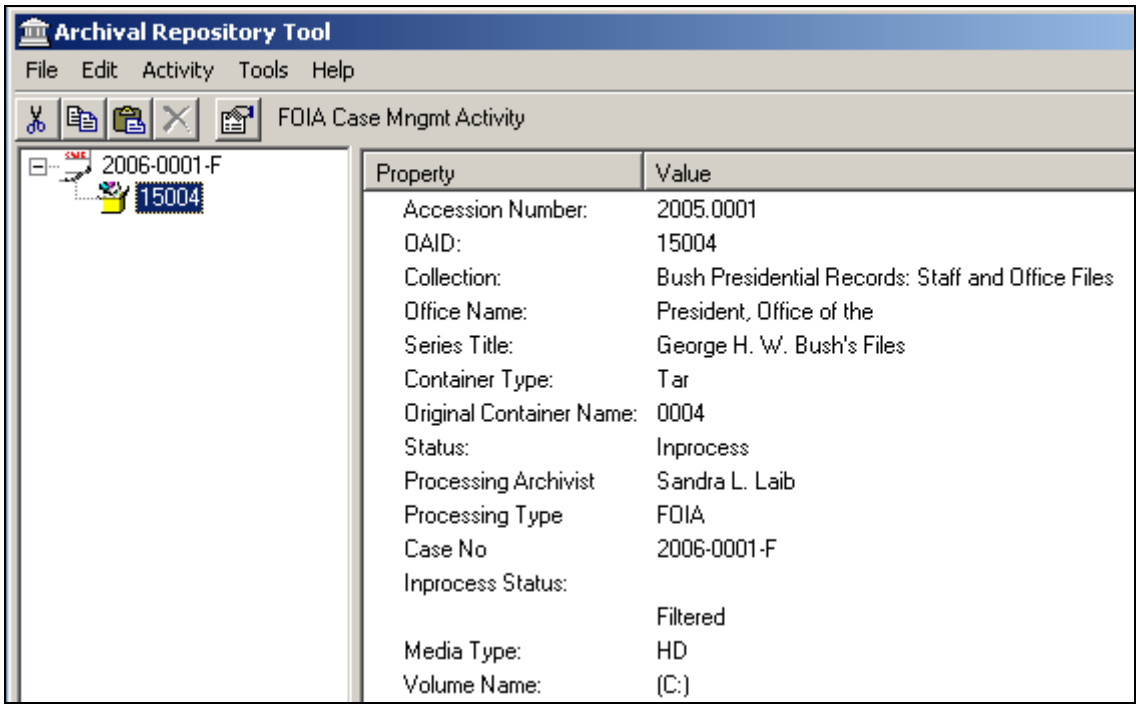
[8] There may be cases where the archivist who created the FOIA case is not the same one who processes or completes the processing of the FOIA. A user with administrative privileges, for example, the Supervisory Archivist, can check a container back in. An option should be provided to enable this user to edit the name of the archivist (computer user name) assigned to the FOIA Case.



A dialog box will appear to allow the archivist to indicate where the checked out container should be placed. Recommended practice is to put it in the Work Area directory in a folder titled with the initials or name of the archivist.



When the container is checked out, the container's manifest is modified to allow the Archival Processing Tool to know which records in the container belong to the FOIA case and which archivist has checked it out. Then, a copy of the container with the modified manifest is placed in the location indicated in the dialog box. The original container with its original manifest is kept in the repository as a backup.

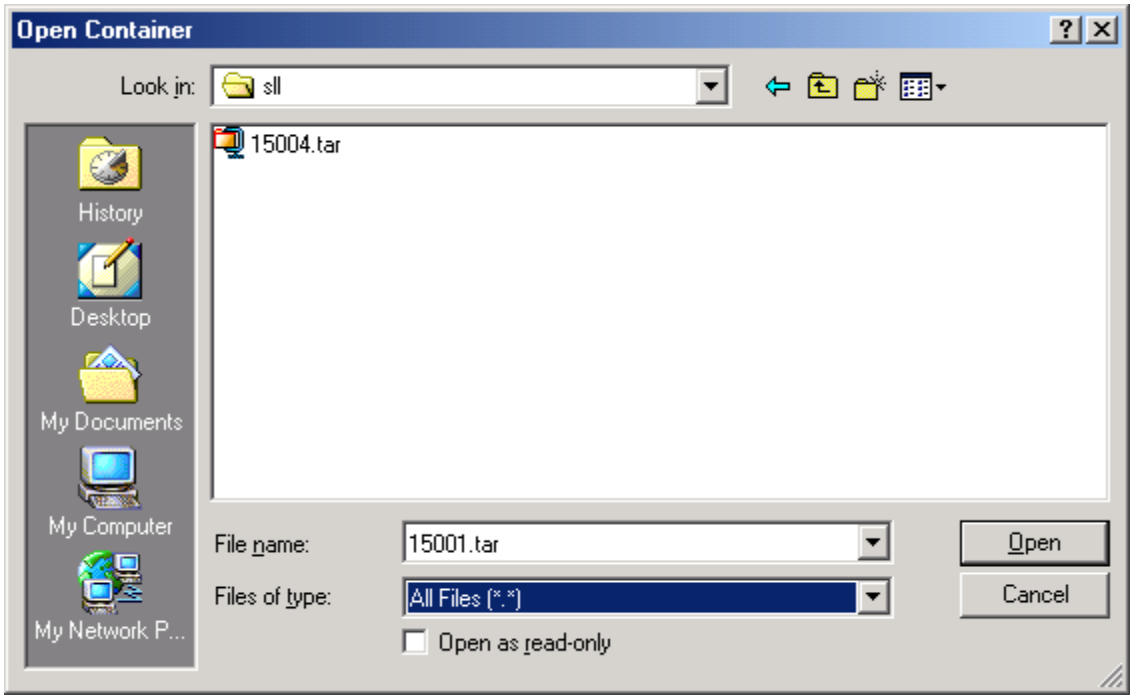


When save is selected, the archivist is returned to the *FOIA Case Mngmt* screen. The *Archivist* who checked out the container, the *Processing Type* under which the container was checked out, and the *Case No* of the case under which it was checked out are displayed in the right pane of the Archival Repository Tool. Containers can only be checked out to one archivist at a time.

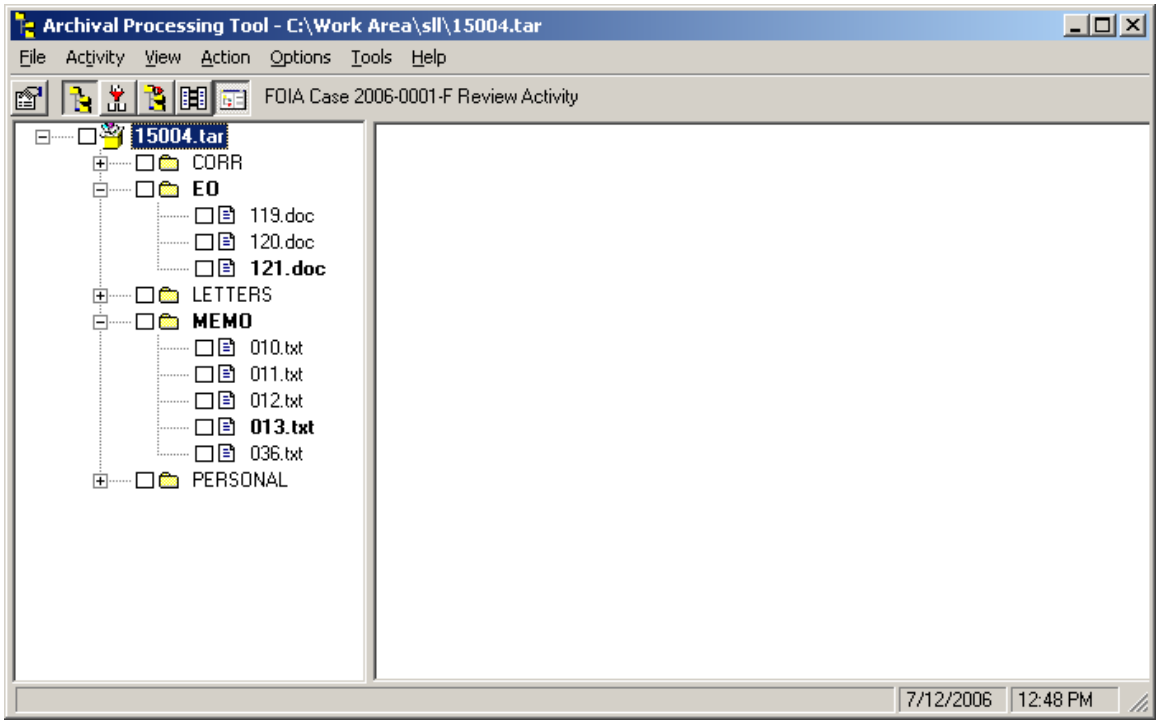
5.2 Open a Container for Review

When a container is checked out, only the archivist who checked out the container or a user with administrator privileges can open the container for processing. Anyone else who opens the container will be opening a read-only copy that remains in the Holdings area.[9] The archivist opens the container using the Archival Processing Tool. The archivist selects *Review* from the *Activity* drop-down menu, and selects *Open* from the *File* drop-down menu.

[9] The archivist who checks out the container can also open the working copy of the container from inside the Archival Repository Tool by selecting *Open* from the *Files* drop down menu.



From the archivist's folder in the Work Area, the container can be opened.



The filenames of the records associated with the FOIA case and the folders that contain the records appear in **boldface**.

5.3 Review Records Related to FOIA Case

To review a record, the archivist opens the file whose file name is boldfaced and it is displayed in a viewer. The archivist reads the document to determine whether it is a non-record, a Personal Record Misfile (PRM), or a Presidential record. Presidential records are reviewed to determine whether there are any PRA restrictions or FOIA exemptions that apply.

The Presidential Records Act defines personal records as follows:

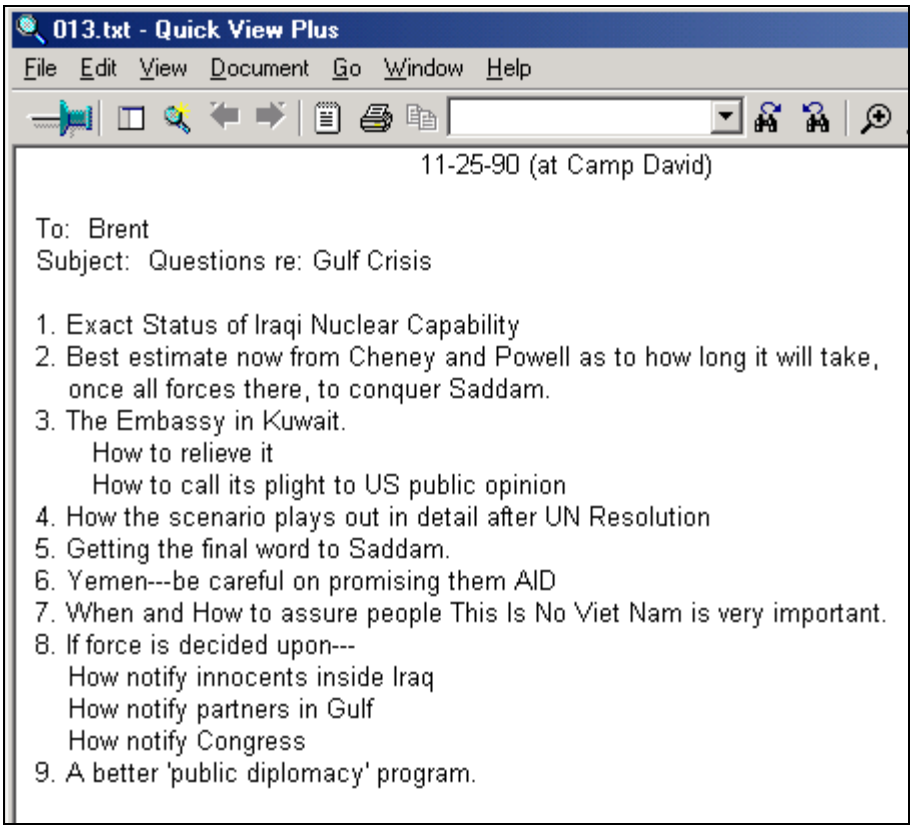
> “The term "personal records" means all documentary materials, or any reasonable segregatable portion thereof, of a purely private or nonpublic character which do not relate to or have any effect upon the carrying out of the constitutional, statutory, or other official or ceremonial duties of the President. Such term includes
>
> a) Diaries, journals, or other personal notes serving as the functional equivalent of a diary or journal, which are not prepared or utilized for, or, circulated or communicated in the course of, transacting Government business.
>
> b) Materials relating to private political associations, and having no relation to or direct effect upon the carrying out of constitutional duties of the President; and
>
> c) Materials relating exclusively to the President’s own election to the office of the Presidency; and materials directly relating to the election of a particular individual or individuals to Federal, State or local office which have no relation to or direct effect upon the carrying out of constitutional, statutory, or other official or ceremonial duties of the President."

A personal record filed with Presidential records is referred to as a Personal Record Misfile (PRM). It should be removed from the Presidential records, and the person who created the record should be notified that the Library is in possession of some of their personal records.

A *Presidential Record* is "documentary materials, or any reasonably segregatable portion thereof, created or received by the President, his immediate staff, or a unit or individual of the Executive Office of the President whose function is to advise and assist the President, in the course of conducting activities which relate to or have an effect upon the carrying out of the constitutional, statutory or other official or ceremonial duties of the President."

Non-records are materials that are neither Presidential Records nor personal/political records. For electronic media, this includes computer operating system files, office application files (word processing, spreadsheets, data base management systems, help files), sample office application files, computer game files, program documentation, and reference e-publications such Webster's Dictionary.

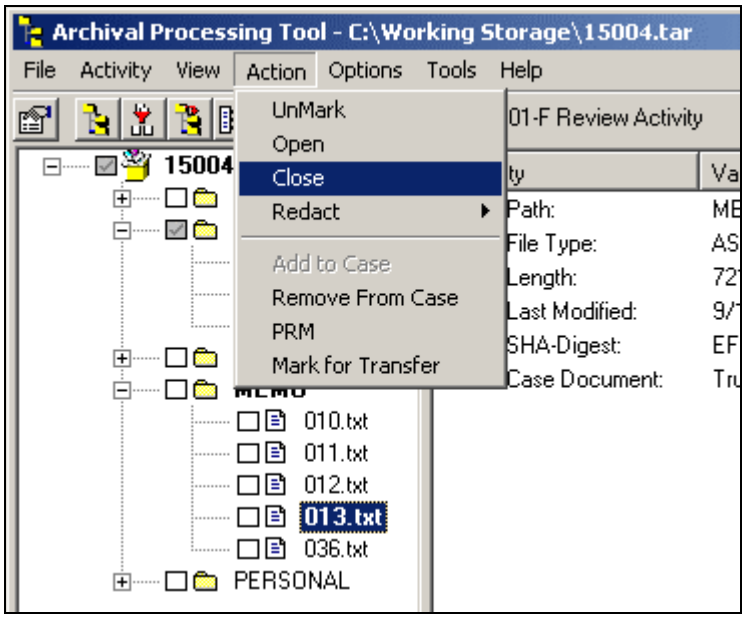
The record shown below is a memorandum from President Bush to Brent Scowcroft, his National Security Advisor. PRA restriction a (5), Confidential Advice, would have applied to this record, had the President not waived that restriction for this record which he included in his book.[10]


013.txt - Quick View Plus
File Edit View Document Go Window Help
11-25-90 (at Camp David)
To: Brent
Subject: Questions re: Gulf Crisis
1. Exact Status of Iraqi Nuclear Capability
2. Best estimate now from Cheney and Powell as to how long it will take, once all forces there, to conquer Saddam.
3. The Embassy in Kuwait.
How to relieve it
How to call its plight to US public opinion
4. How the scenario plays out in detail after UN Resolution
5. Getting the final word to Saddam.
6. Yemen---be careful on promising them AID
7. When and How to assure people This Is No Viet Nam is very important.
8. If force is decided upon---
How notify innocents inside Iraq
How notify partners in Gulf
How notify Congress
9. A better 'public diplomacy' program.

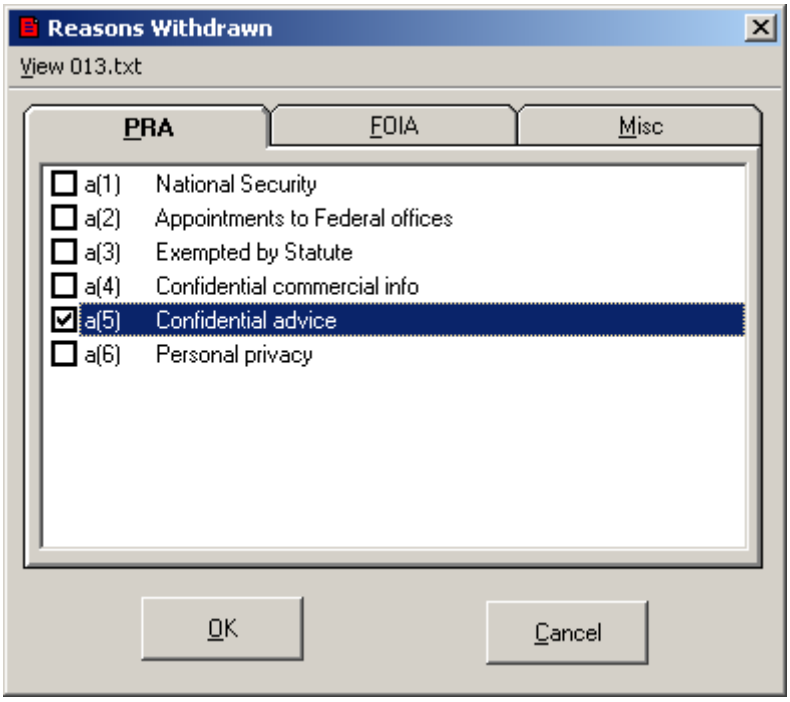
By selecting an action from the Action drop-down menu, an archivist can Open, Close, or Redact the record, or Mark it as a Personal Record Misfile or Mark it for Transfer to the Library (because it is a non-record).[11] Suppose that an archivist decides this record should be closed because of PRA a (5). He selects *Close* from the Action drop-down menu.

[10] *All the Best, George Bush: My life in letters and other writings*, 1999, pp 490-491.

[11] It might be useful to add a review action for situations when the archivist needs further guidance before making a review decision; something like "Pending" or "To Be Determined" or "Needs Decision" that will highlight that the review action of the record is outstanding.



If a record is closed, the archivist must indicate the reason for withdrawal, in this case PRA a (5). There are no FOIA exemptions for this particular record.[12] [13] [14]

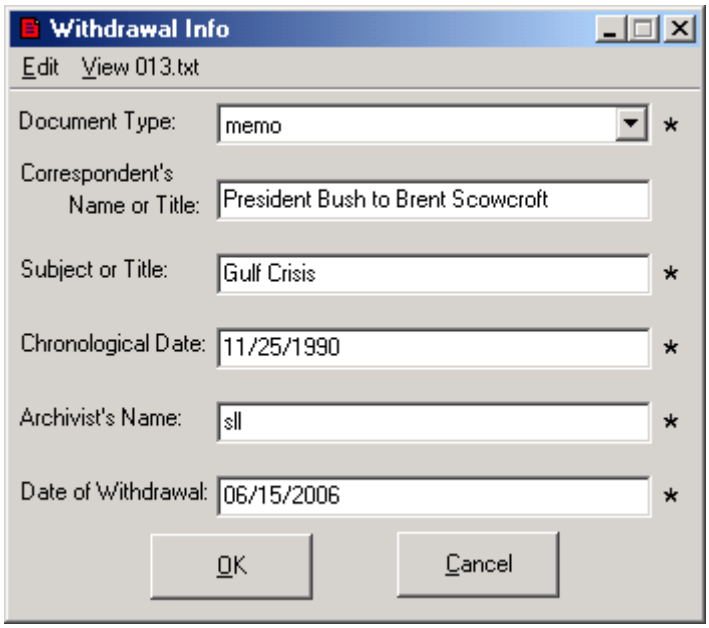


[12] It might be useful for the Reasons Withdrawn window to have a minimize function, as the archivist may need to look at the record in order to determine all of the reasons withdrawn.

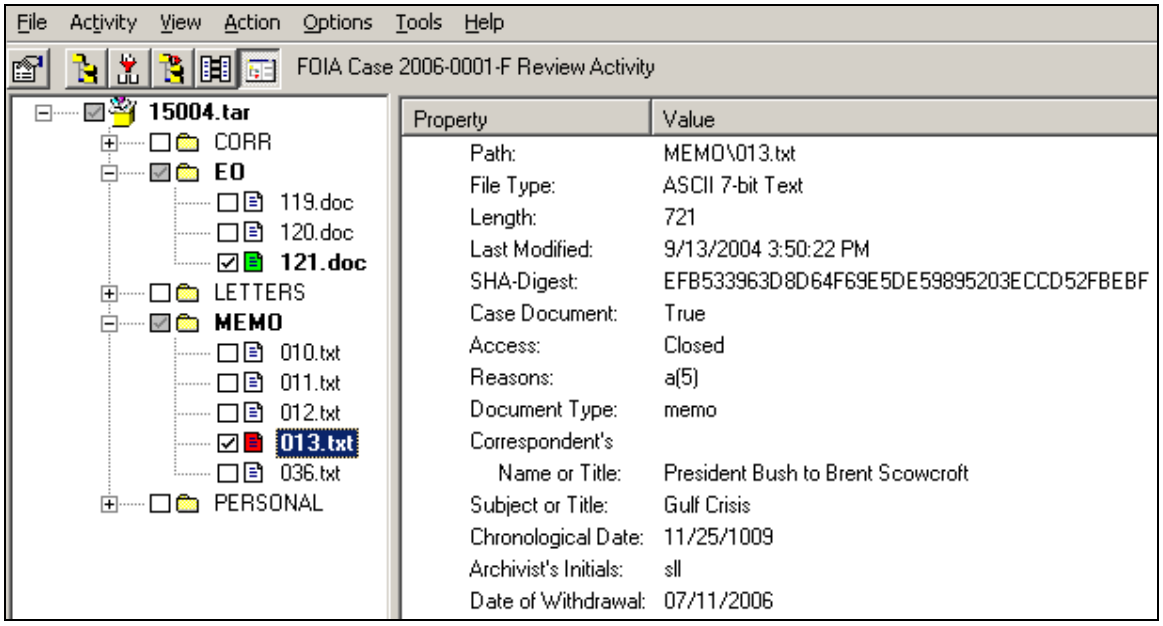
[13] FOIA exemption b(7), Law Enforcement Investigations, allows selection of subcategories of that exemption.

[14] Reasons for Withdrawal a(3) b(3), Exempted by Statute, should enable selection or entry of the relevant statute(s). One option would be to have a list of statutes for which there are exemptions to release and when a(3) b(3) was selected, the archivist was required to pick one or more statutes from that list. The list could be extended by the capability to Edit Reasons Withdrawn.

The archivist must also enter withdrawal information—record type, correspondent's name(s), subject or title, and chronological date. Some of this information can be copied from the record and pasted into the withdrawal form. The Archivist's username is automatically captured, as is the date of withdrawal.



A box next to the filename of the record is color coded to indicate the type of access—green for open, red for closed, blue for redacted. Yellow indicates a PRM and grey indicates marked for transfer. The access type, access restrictions, and withdrawal information for a record are displayed in the right panel of the window and are stored in a manifest file in the container.



A check in a grey box next to a folder (or container) symbol indicates that some of its records have been reviewed. When all of the records in a folder (or container) have been reviewed, the check will appear in a white box.
If a record is redacted, a document image is created for the record, and the archivist can block out text in the image, select the reason for withdrawal and stamp an area near the text with the reason for withdrawal.

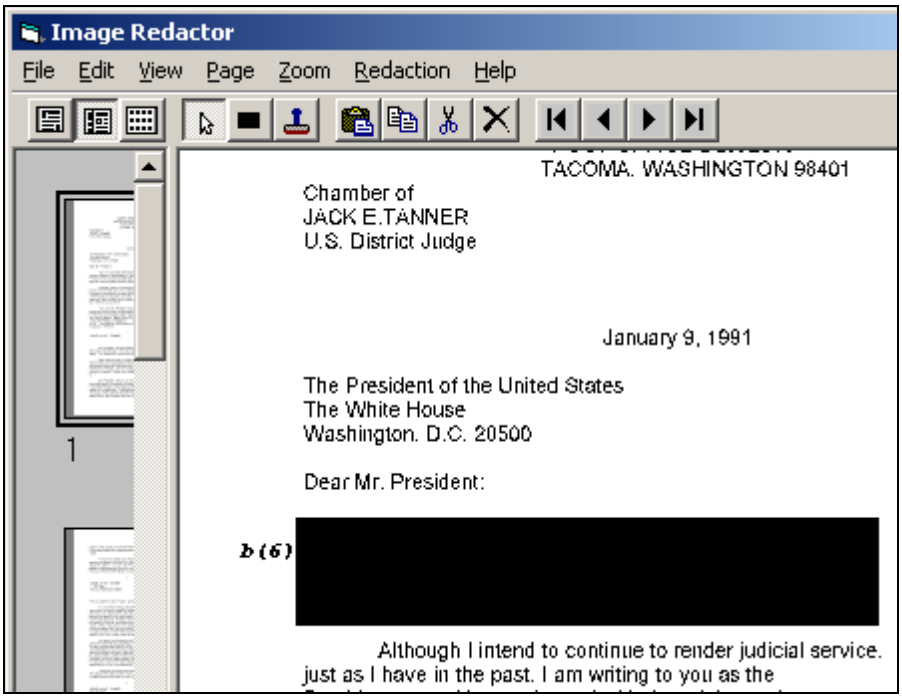


Image Redactor
File Edit View Page Zoom Redaction Help
TACOMA. WASHINGTON 98401
Chamber of
JACK E.TANNER
U.S. District Judge
January 9, 1991
The President of the United States
The White House
Washington. D.C. 20500
Dear Mr. President:
b(6)
Although I intend to continue to render judicial service.
just as I have in the past. I am writing to you as the

An archivist may make a review decision and later determine that it was an incorrect decision. This decision can be reversed by highlighting the filename of the opened, closed, or redacted file, or the filename of a file marked PRM or marked for transfer, and then selecting Unmark from the Action drop-down menu.

An archivist may update (rather than reverse) a previous decision. Examples include when the record has been declassified, opened on appeal, and at the end of the 12-year PRA restriction period (when only FOIA exemptions stand). Support for these capabilities are discussed in the Review Section of the PERPOS Reference Manual [Underwood et al 2006].

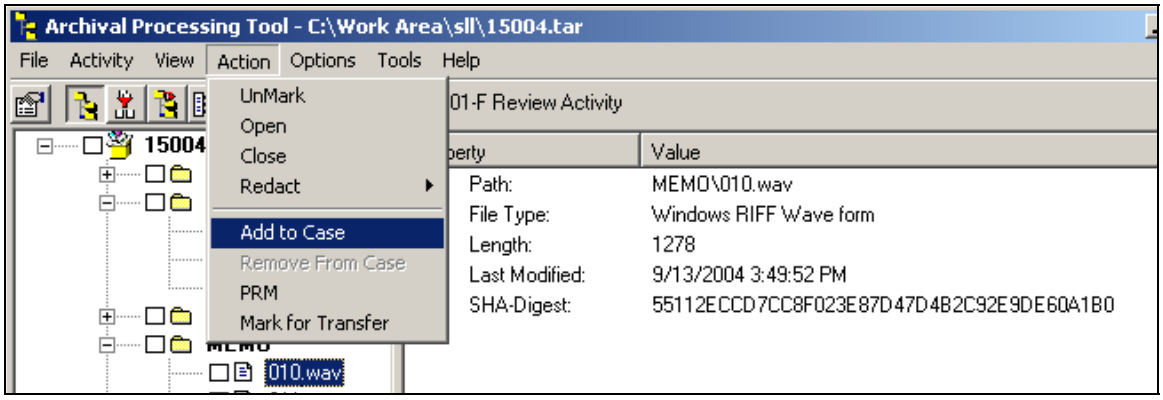
The review support capabilities are essentially the same as were developed for systematic processing, except that only the records relevant to the FOIA query need to be reviewed. The support capabilities for review of records are described detail in the PERPOS User's Reference Manual [Underwood et al 2006].

5.4 Add to or Remove Record from a FOIA Case

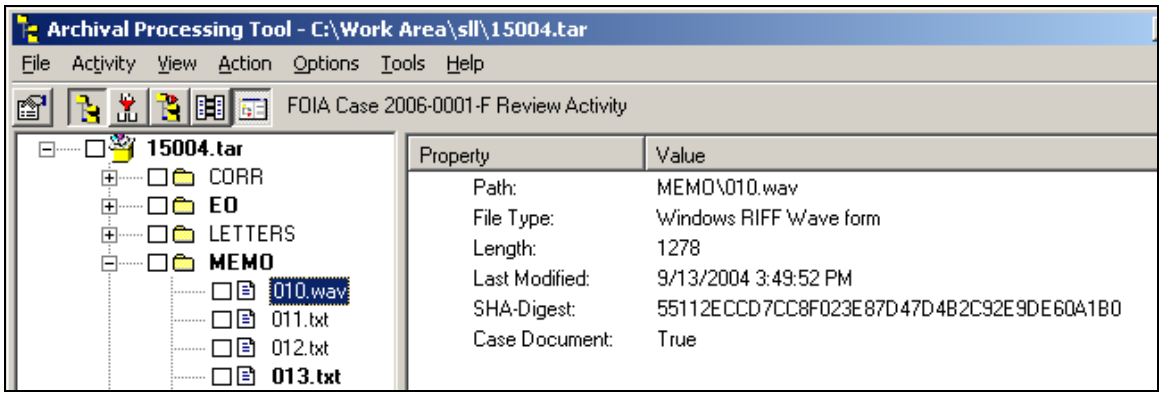
Two additional actions have been added to the *Action* drop down menu in the *Review Activity*. These actions are *Add to Case* and *Remove from Case.* It is possible that some records were not returned in the result set of a FOIA search because the records were not indexed. This situation can occur when files have not been extracted from an archive file, a file is password protected or is an image or audio file. It is also possible that a record read by the reviewer could be determined to be relevant to the FOIA request, but was not relevant to the FOIA query.

5.4.1 Add a Record to a FOIA Case

While viewing files in a container that is related to a FOIA case, an archivist can view files that are not in the result list of the FOIA case. If after viewing a file that is not in the results list of the FOIA case, an archivist decides that the record is relevant to the FOIA case, he can add the file to the results list of the current FOIA case container by highlighting the filename and selecting *Add to Case* from the *Action* drop-down menu.

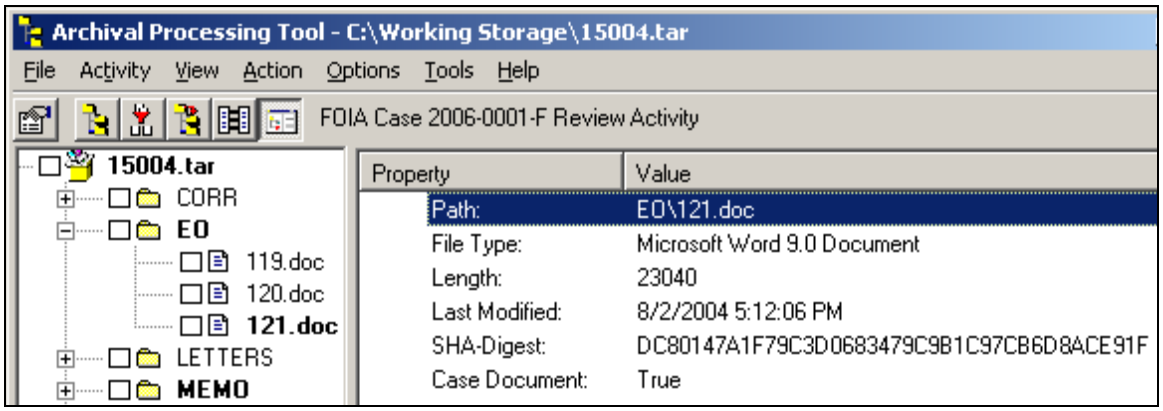


The Case Document property will be set true for the file and the filename will appear in boldface indicating that it is a member of the FOIA case. If there are any other versions of the record, e.g., redacted or in a different file format, they will have the same GroupID, and they will appear in boldface.

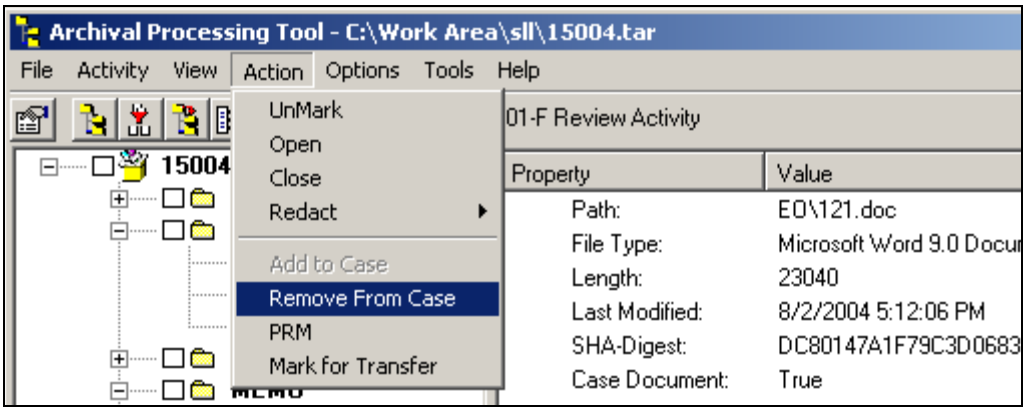


5.4.2 Remove a Record from a FOIA Case

After viewing a record, an archivist may decide to remove it from the result list of the FOIA case. This might occur because the file was an operating system or software application file, program documentation or a sample text file that should have been filtered and transferred to the Library.[15]
The following figure shows record "EO\121.doc" with the file name and the path boldfaced and the *Case Document* property set "True."

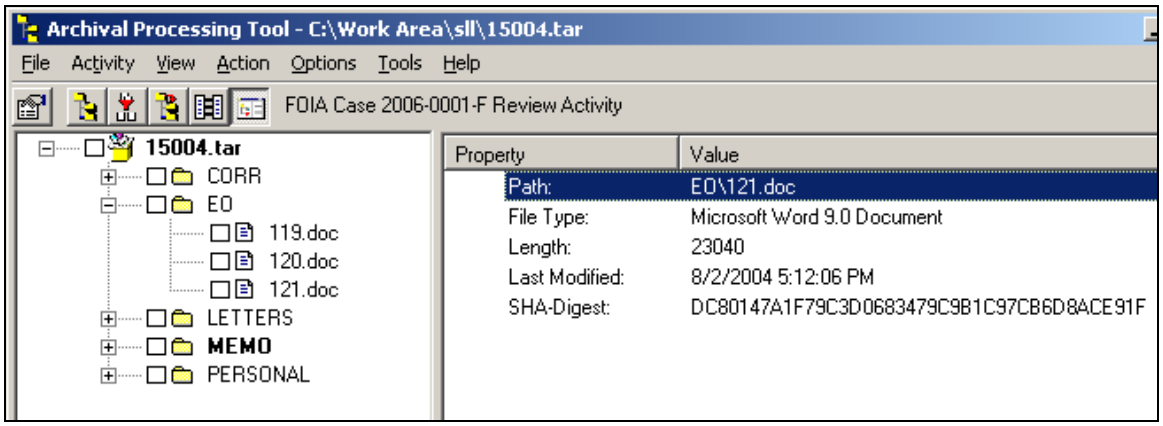


To remove this record from the FOIA Case, the archivist first highlights the filename and the selects *Remove from Case* from the *Action* drop-down menu.



The Case Document property will be set false, the file name will not appear in boldface, and the Case Document property will not be displayed. If there were any other versions of the record, e.g., redacted or in a different file format, they will have the same GroupID, and their boldface will disappear.

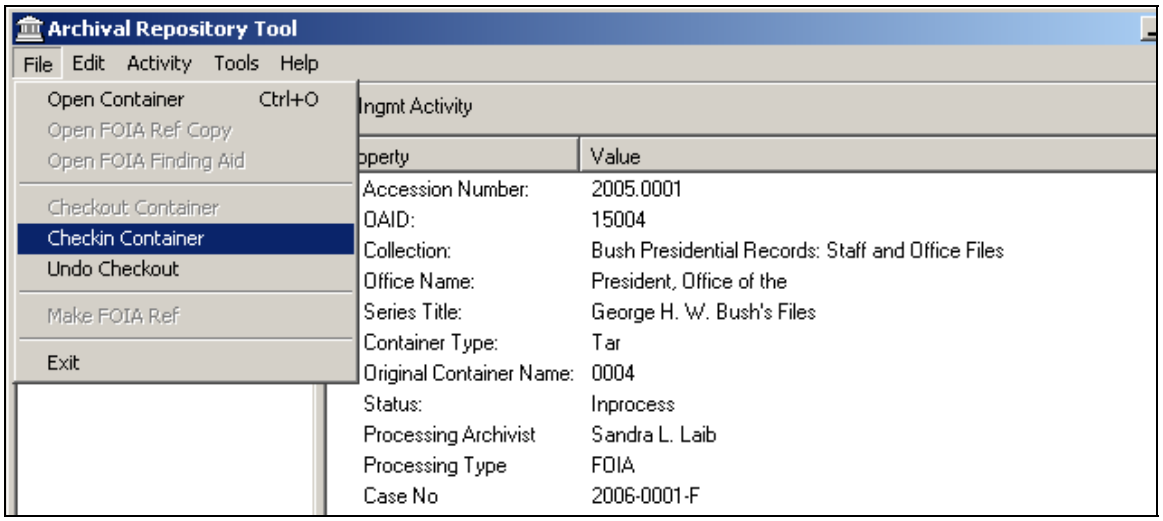
[15] This function may not be needed. If the only reasons for removing a file from a case are that it is a non-record or a PRM, then the archivist should mark it for transfer or as a PRM. Having done so, the result list is not modified, but when the FOIA reference copy is created, it will not contain files that are in the result list that are marked as PRMs or are Marked for Transfer.



During review, when there is both an original record and a copy of the record that has been converted to a different file format, if either record is marked closed, both records are marked closed and given the same reason for closure. When any record in the group is marked Open, PRM, or for Transfer, then all copies of records with the same GroupID are marked with the same type of access. When a record is redacted, both the original and any copy converted to a different file format are marked closed.

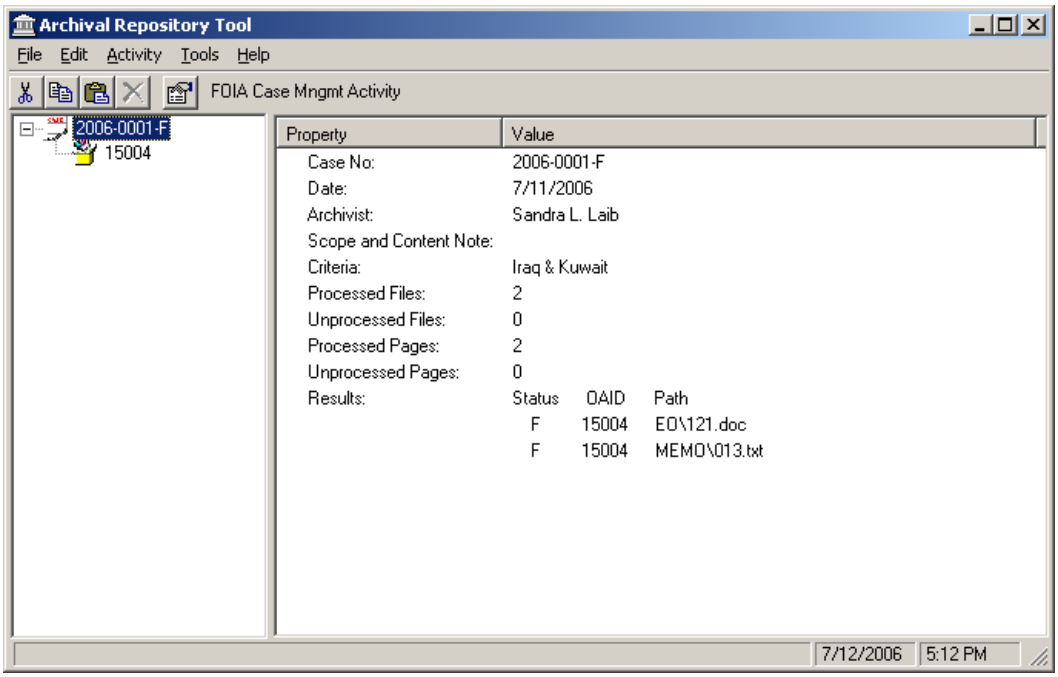
5.5 Check in a Container

When an archivist has finished reviewing all the records in a container that are part of a FOIA Case, he checks the partially reviewed container back into the archival repository (including both reviewed and un-reviewed records). This is accomplished in the Archival Repository Tool by selecting *Checkin Container* from the *Files* drop-down menu.



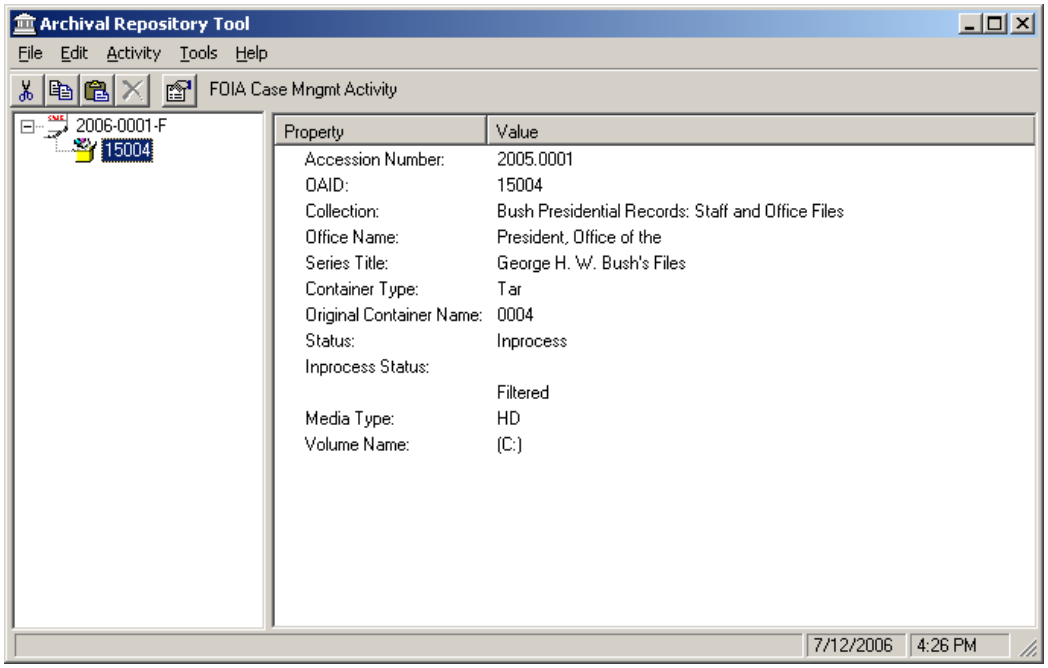
When a container is checked back into the repository, the FOIA case under which the container was checked out and all other FOIA cases that contain one or more of the same records are updated. Their *Processed Files* and *Unprocessed Files* properties are updated. The Case's *Processed Pages* and *Unprocessed Pages* properties are not automatically

updated. It is the responsibility of the archivist to update these properties by selecting the *Estimate Pages* tool. The *Status* field, which displays Results, is also updated for this case and any other case that contains the reviewed records. Any additional records, created during redaction or conversion, are added and any path properties that were changed due to changes in arrangement are modified. Temporary information (Archivist's name and results list for this container) is removed from the manifest and the reviewed container and manifest replaces the container and manifest in holdings.



When a record is removed from a FOIA case, all other records with the same GroupID are removed from that case, but no records are removed from any of the other cases. The records removed from one FOIA case are still relevant to other FOIA cases to which they belong.

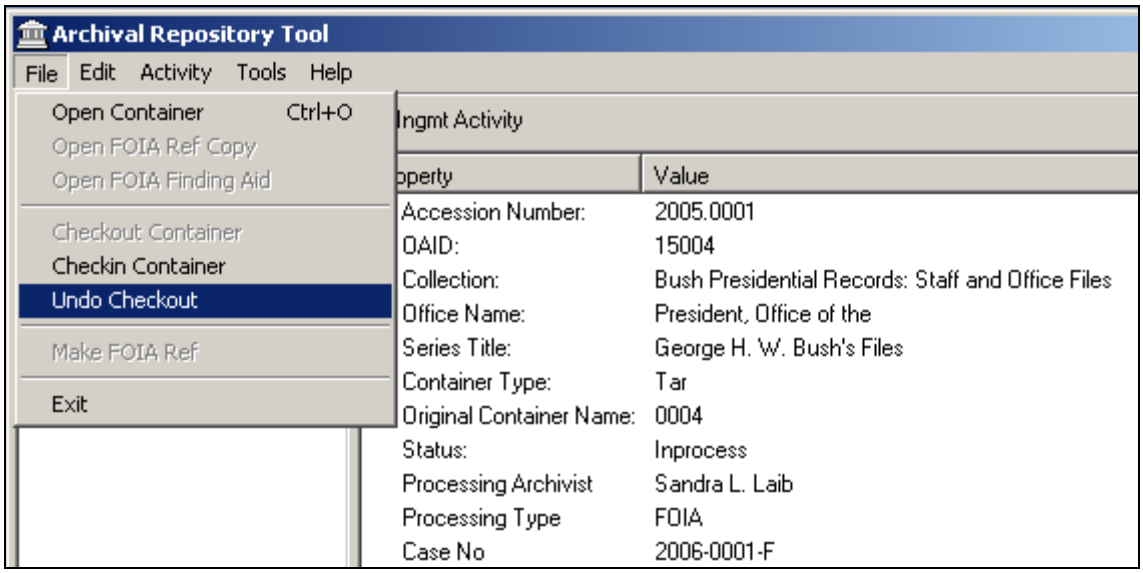
Once all pertinent updates have been made, the temporary information that was added to the container's manifest is removed and the container replaces the original container that had been kept as a backup in the repository.



When an archivist has processed a container, he can proceed to the next container that is not checked out to another archivist. All the relevant records in some of the containers may have already been processed. If this is the case, the container does not need to be checked out. The archivist can tell if a container has any unprocessed records associated with the FOIA case by looking at the FOIA case's result list. The status of a record is listed beside the filename of the record. They can stop review at any time and resume review later by selecting the relevant FOIA case number from the Archival Repository Tool.

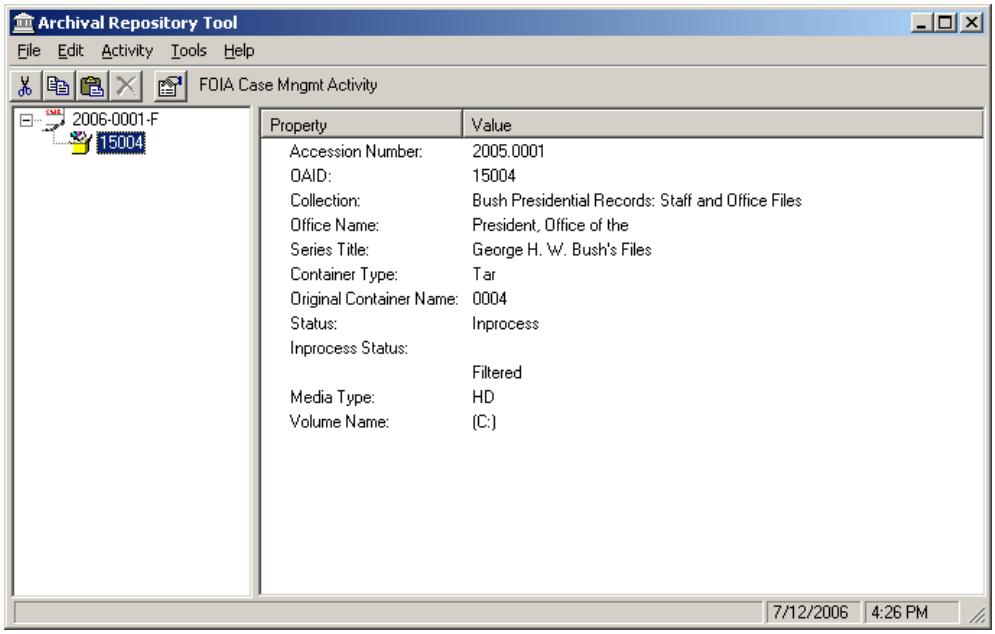
5.6 Undo Checkout

The working copy of a container may become corrupted or be inadvertently deleted. In this case, the archivist does not want to, or cannot, check the container back into the repository. In fact, the archivist should undo the whole check out process and start over. This can be accomplished in the Archival Repository Tool by selecting *Undo Checkout* from the *Files* drop-down menu.



There is a check that the archivist who is undoing the check out is the archivist who checked out the container. The container reverts to the copy that remained in holdings. It is the responsibility of the archivist who undoes the check out to delete the copy in his work area.[16]

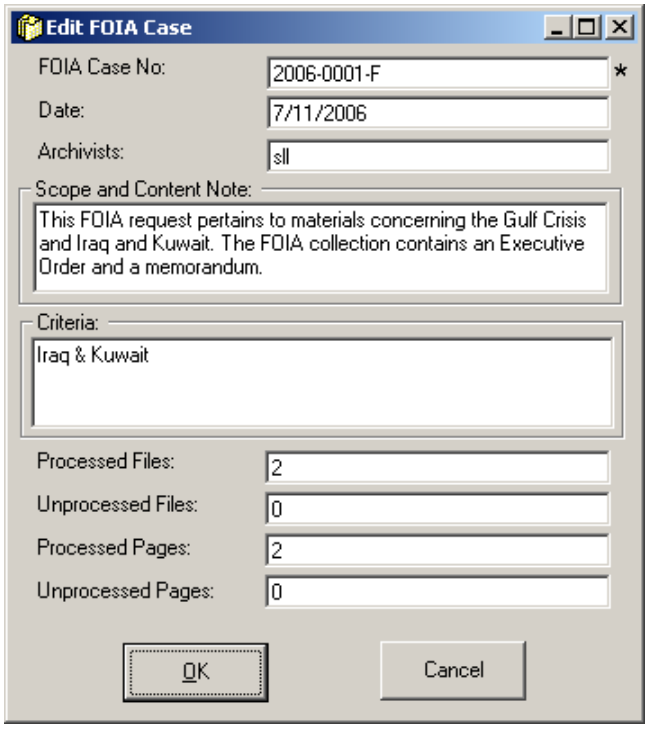
The *Archivist*, *Processing Type*, and *Case No* are removed from the container's property list and will not be displayed in the right pane of the Archival Repository Tool and the container becomes available for checkout once more.



[16] Searching the work area for the name of the container could eliminate this secondary responsibility. If it was found, then it could be deleted. If not found, it had been inadvertently deleted.

6. FOIA Case Description

When an archivist has reviewed the records in all the containers and folders that are associated with a FOIA Case, the FOIA Case should be described. To describe the FOIA Case an entry is made in the FOIA Case's *Scope and Content Note* property. To accomplish this, select the FOIA Case in the *FOIA CASE Mngmnt* Activity, and then select *Properties* from the *Edit* drop-down menu.



Once the *Scope and Content Note* has been entered, select the *OK* button.[17] The archivist will be returned to the *FOIA Case Mngmnt Activity* screen. The *Scope and Content Note* will be displayed with the other properties in the right pane. After a description has been entered into the *Scope and Content Note* field, the *Make FOIA Ref* menu option on the *File* drop-down menu will be enabled.

7. FOIA Reference Container, its Arrangement and Finding Aid

The last step in FOIA Processing is creating the FOIA Collection Reference Container and Finding Aid. The Archival Repository Tool creates them at the same time. The arrangement of the information in the Finding Aid is the same as the arrangement of the

[17] The Finding Aid should specify that these are electronic, not paper records.

records in the Reference Container. In addition, the information at the beginning of the Finding Aid is the same as the information in the Manifest header.

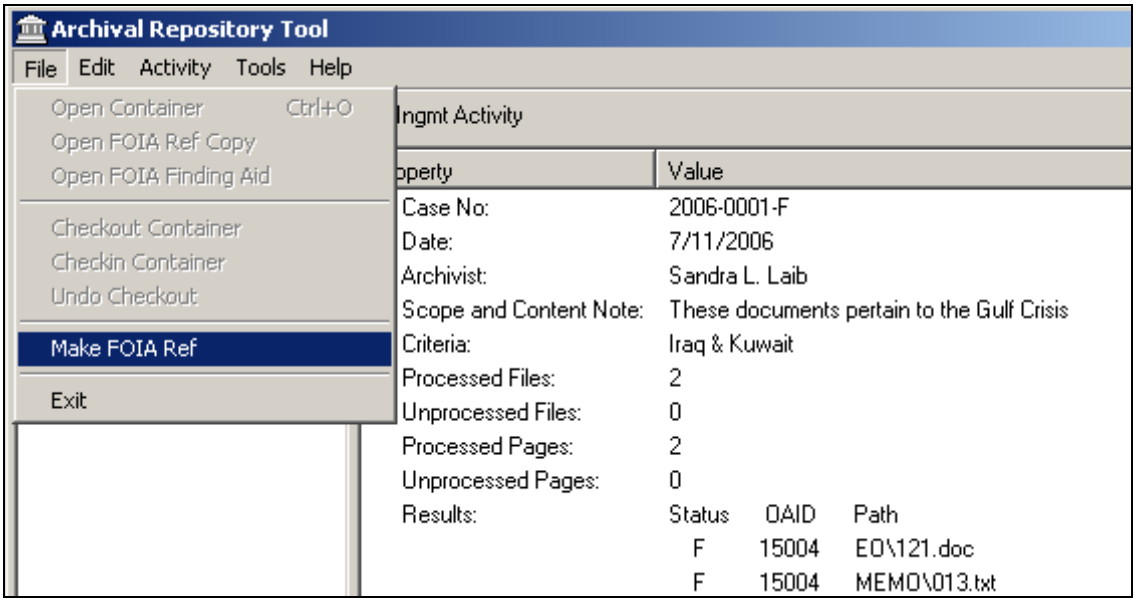
7.1 FOIA Reference Container and Finding Aid Arrangement

The contents of the FOIA Case Reference container and the Finding Aid are arranged following the Bush Presidential Library guidelines for arranging a FOIA Collection. The FOIA Collection Reference container is arranged by Collection with the *Bush Presidential Records: Staff Member and Office Files* being the first collection. Under each collection, the offices are arranged alphabetically. Under each office, the series are arranged alphabetically by the staff member's last name. Under each Series, the containers are arranged numerically by OA/ID number (smallest to largest). Within OAID number, the directories and records are in the order in which they appeared in the original container, unless there was some rearrangement. The collection of *Quayle Vice Presidential Records* are arranged next in alphabetical order by office, then alphabetically by staff member name, and then in OA/ID number order (See example in Appendix C).

If no staff person is identifiable for a series, the series are arranged alphabetically after the last staff person for an office. If Federal records are processed in response to a FOIA request, they are arranged by Record Group at the very end of the FOIA Collection.

7.2 Make a FOIA Reference Container and Finding Aid

To create a FOIA Reference container select the FOIA Case in the *FOIA CASE Mngmnt* Activity, then select *Make FOIA Ref* from the *File* drop-down menu. If the *Make FOIA Ref* menu option is not enabled, then not all files in the result set have been reviewed or the *Scope and Content Note* for the FOIA Case has not been created.

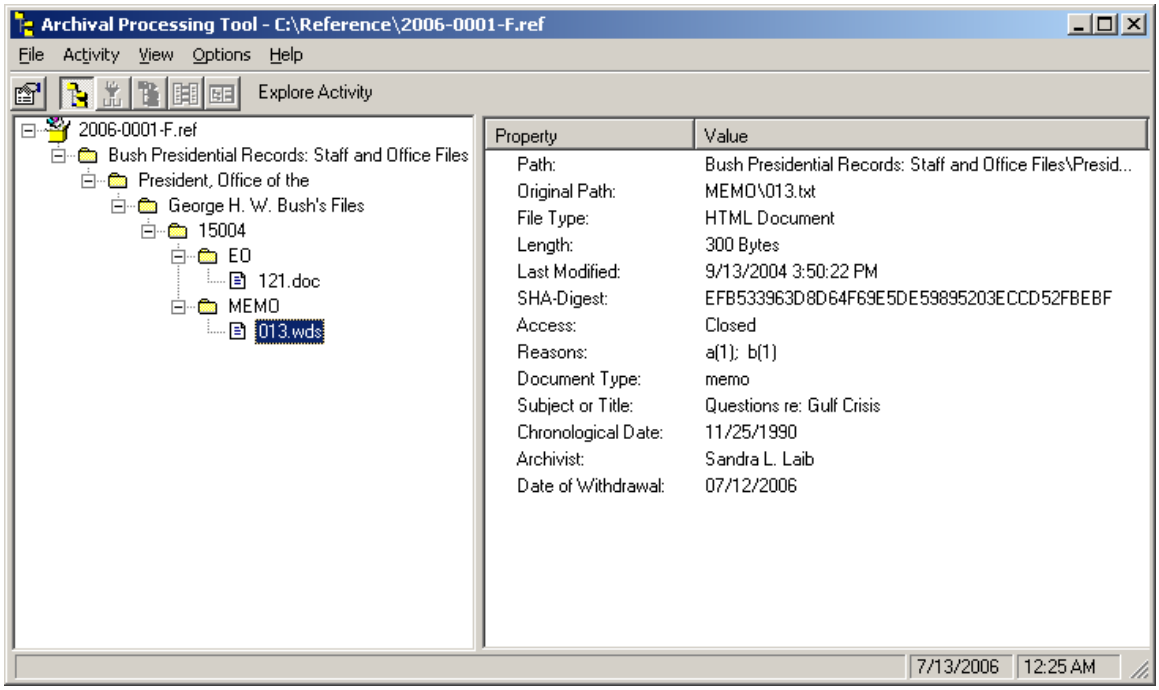


The FOIA Reference Container will contain opened records, redacted records, FOIA markers for relevant records that are in another FOIA Collection or in a Reference container for a systematically processed record series, withdrawal sheets and transfer sheets. It will not contain closed records, originals of redacted records, PRMs, or transferred records.

When the FOIA Case is selected, the case containers are shown under the FOIA Case and the contents of the Scope and Content Note along with query and the result list are displayed in the right pane. The only difference is that the *Open FOIA Ref Copy* and the *Open FOIA Finding Aid* option of the *File* drop-down menu are enabled when a FOIA Case that has a Reference Container is highlighted.

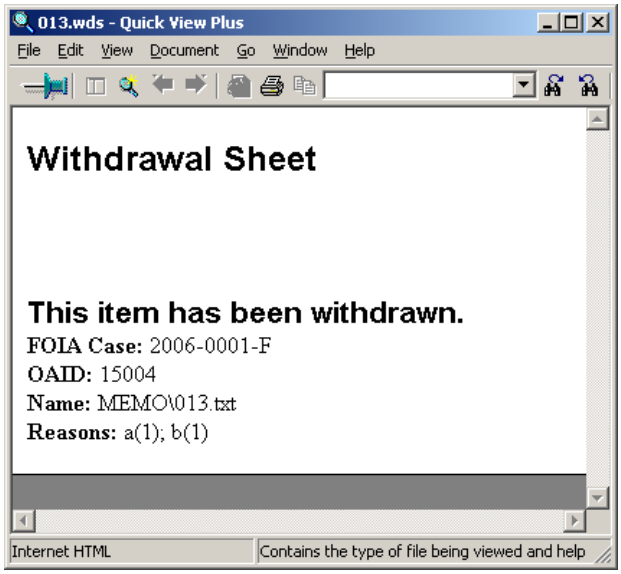
7.3 Open a FOIA Reference Container

To open the FOIA Reference container, select the FOIA Case. Then select *Open FOIA Ref Copy* from the *File* drop-down menu. This will open the FOIA Reference Container inside the Archival Processing Tool.



The "wds" extension on the highlighted filename indicates that it is a "Withdrawal Sheet." In right pane, the *Access* property is shown as "Closed." The grayed out Activity buttons on the tool bar indicate that the container is read only. To view the Withdrawal Sheet, select *QuickView* from the drop-down *View* menu.[18]

[18] The withdrawal information shown in the right windowpane (Document Type, Subject or Title, Chronological Date, Creator name and Office) should have also been shown on the withdrawal sheet.


013.wds - Quick View Plus
File Edit View Document Go Window Help
Withdrawal Sheet
This item has been withdrawn.
FOIA Case: 2006-0001-F
OAID: 15004
Name: MEMO\013.txt
Reasons: a(1); b(1)
Internet HTML
Contains the type of file being viewed and help

The figure below shows the kinds of metadata in the Manifest of a FOIA Reference container.

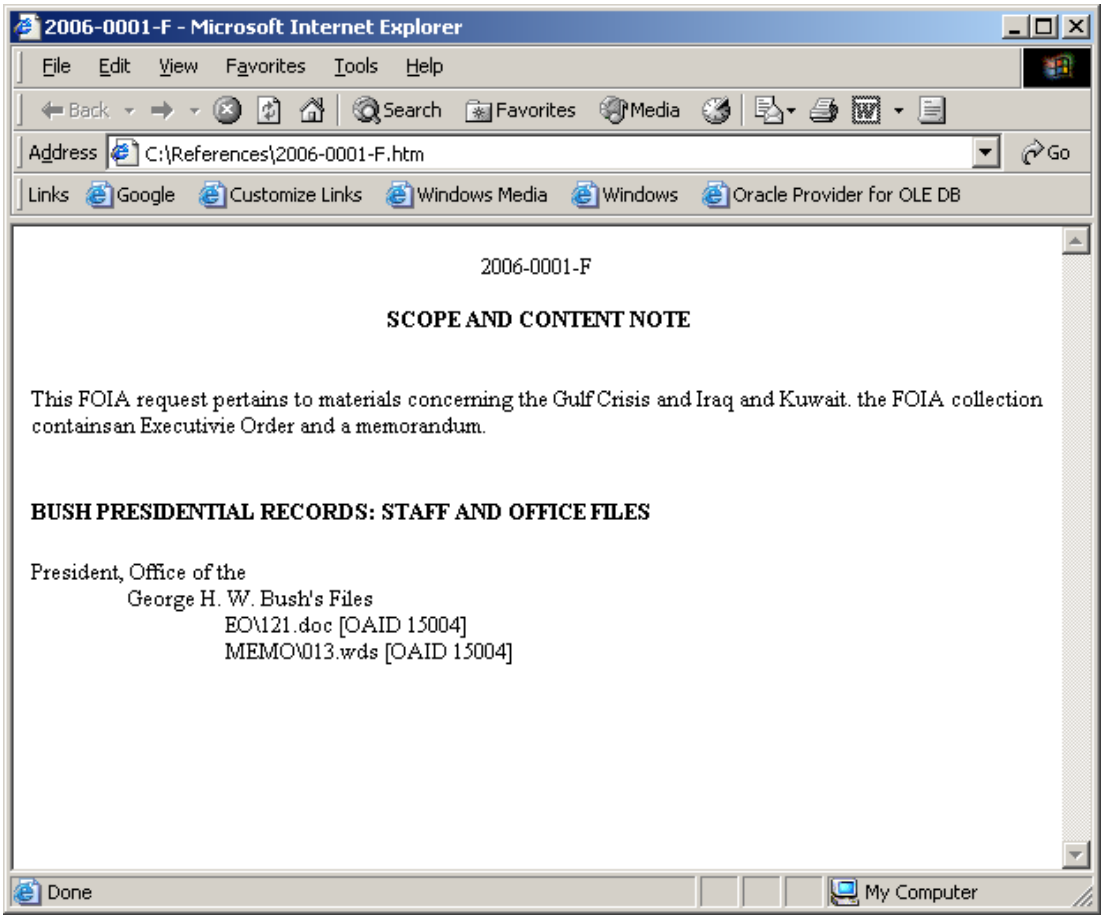
```
Manifest-Version: 5.0
FOIA ID: 2006-0001-F
Scope and Content Note: This FOIA request pertains to materials concerning the Gulf Crisis
and Iraq and Kuwait. the FOIA collection containsan Executivie Order and a memorandum.

Collection: Bush Presidential Records: Staff and Office Files

Office: President, Office of the

Series: George H. W. Bush's Files

OAID: 15004

Name: 15004\
ItemID: 1
FileType: Directory

Name: 15004\EO\
ItemID: 2
FileType: Directory

Name: 15004\EO\121.doc
ItemID: 3
SHA-Digest: DC80147A1F79C3D0683479C9B1C97CB6D8ACE91F
OriginalPath: EO\121.doc
FileType: Microsoft Word 9.0 Document
Access: Open
ProcessedType: F

Name: 15004\MEMO\
ItemID: 4
FileType: Directory

Name: 15004\MEMO\013.wds
ItemID: 5
SHA-Digest: EFB533963D8D64F69E5DE59895203ECCD52FBEBF
OriginalPath: MEMO\013.txt
FileType: HTML Document
Access: Closed
ProcessedType: F
```

The Manifest indicates the folder titles, and then file metadata, any Withdrawal sheets that are included in the container in place of closed records, and FOIA Markers for records that are part of another FOIA Reference container or in a Reference container of a systematically processed record series.

7.4 Open a FOIA Collection Finding Aid

While reviewing records in a container (or directory), an archivist may find there are additional records not in the result set that if reviewed would compete the review of the container (or directory). If reviewed at that time, these additional records are said to be incidentally processed. The Finding Aid for that FOIA Collection of paper records will indicate records that are included that are not relevant to the FOIA request, but were incidentally processed. Currently, the FOIA Reference Container and the Finding Aid do not include records that were incidentally processed.[19]

To open the Finding Aid of a FOIA Case, first select the FOIA Case. Then select *Open Finding Aid* from the *File* drop-down menu.


2006-0001-F - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address C:\References\2006-0001-F.htm
Go
Links Google Customize Links Windows Media Windows Oracle Provider for OLE DB
2006-0001-F
SCOPE AND CONTENT NOTE
This FOIA request pertains to materials concerning the Gulf Crisis and Iraq and Kuwait. the FOIA collection containsan Executivie Order and a memorandum.
BUSH PRESIDENTIAL RECORDS: STAFF AND OFFICE FILES
President, Office of the
George H. W. Bush's Files
EO\121.doc [OAID 15004]
MEMO\013.wds [OAID 15004]
Done
My Computer

[19] It is possible for an Archivist to review records whose filename was not in boldface. The APT just does not keep track of the fact that they were incidentally processed. However, this option can be added.

The Finding Aid for the FOIA Case is created as an HTML document so that it can be published on the Bush Presidential Library Web Site.

8. Inactive FOIA Case Files

After a requestor has viewed the materials of a FOIA collection of paper records, or the case is relatively inactive, that is, other researchers are not referring to it, an archives technician breaks apart the FOIA case of paper records and files the folders comprising the FOIA Collection back into the stacks shadow file. This is not necessary for electronic records in the Archival Repository as all the records of a series are in the Master containers.

When a FOIA collection of paper records is broken down, folders that are referenced by other FOIA cases are moved into one of the other cases referencing it. If we simply delete the inactive FOIA Reference Collection, we need to decide what to do about the other FOIA collections that reference records in the inactive (to be deleted) FOIA collection. One way to handle the problem would be to unmark the FOIARef field in the ContainerDocuments table for all records that had the inactive FOIA Case No in the field. After that, all the other FOIA Collection Containers that referenced a record in that inactive collection would need new containers created. One of the other FOIA Cases would then get a copy of the record and the other would reference the updated FOIA Collection. This would have to be done for each record that was referenced by other FOIA Cases when one was destroyed. This could become very complicated. A single FOIA Case Container could have to be remade several times.

The stacks shadow file of paper records gets a FOIA Marker for records that are in a FOIA Case when systematic processing in performed after some FOIA processing has been done. Currently, the Archival Repository Tool is not handling this situation the same way. When a systematic reference copy of a container is made, it gets copies of all open and redacted files. Even if copies of some of those files are in FOIA reference containers, the systematic reference container gets a copy of the file, not a FOIA Marker.

There is a simpler way to handle the situation, which would make handling inactive FOIA cases simpler. It would also make creating the FOIA Reference Container simpler. For electronic records, the original record stays in a master container, and the systematic reference container gets a copy of the record. Why not give all FOIA cases that reference a record a copy of the record? It would be easier for the researcher if every record that pertains to his request were in a single container. He would no longer get some of his requested records and have to refer to other FOIA collections to see the remainder.

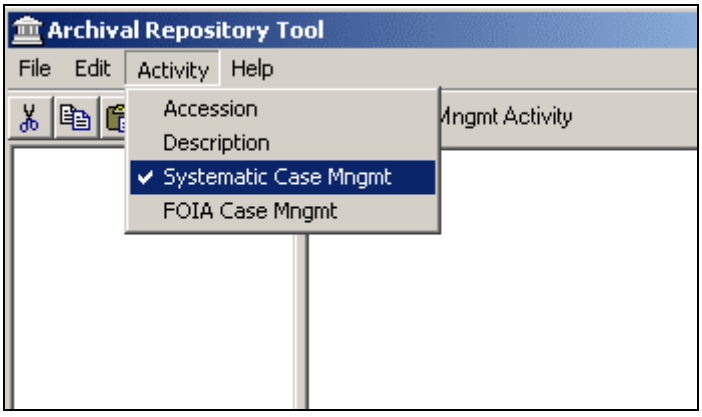
9. Changes to Systematic Processing

A *Systemic Case Mngmt* activity has been added to the Archival Repository Tool. This was necessary so that a distinction could be made between FOIA Processing and Systematic Processing. This also makes it possible to keep track of which archivist is responsible for systematically processing specific containers. It also makes it possible for an archivist to know which kind of processing is being performed on a container that has been checked out by another archivist.

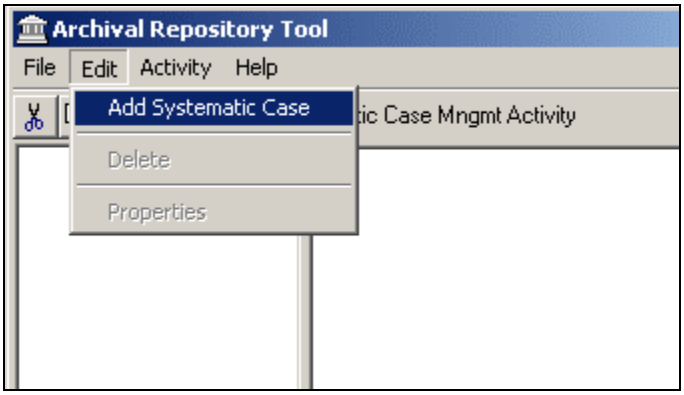
It is now necessary to create a Systematic Case before any systematic processing can be performed on a container. The container must be added to an existing systematic case before the container can be checked out for systematic processing.

9.1 Creating a Systematic Case

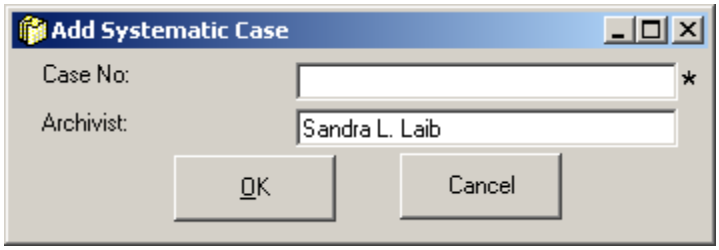
The Archival Repository Tool is used to create a Systematic case. Select *Systematic Case Mgmnt* from the *Activity* drop-down menu.



Select *Add Systematic Case* from the *Edit* drop-down menu.



This causes the Add Systematic Case dialog box to appear.



The Archival Repository Tool automatically fills in the archivist's username as the *Archivist* name. The format of the Systematic Case No is

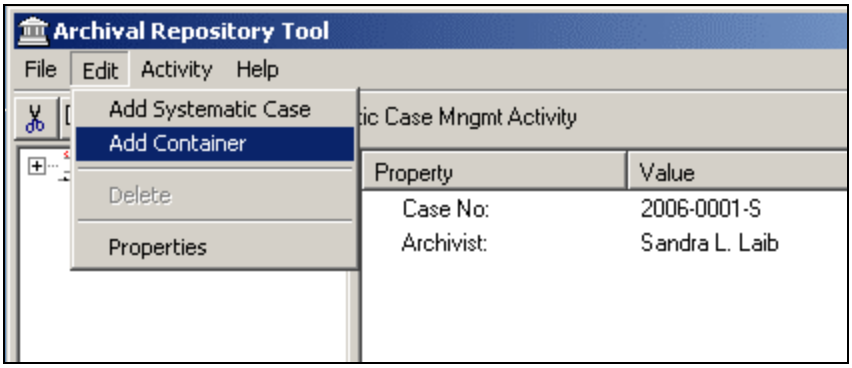
YYYY-XXXX-S

where YYYY is a 4-digit year, XXXX is a sequential number beginning with 0001, and S indicates a systematic case as opposed to a FOIA case. The archivist enters the Systematic Case number then selects *OK*. There is a check to be sure that there is not a collision with an existing systematic case number.

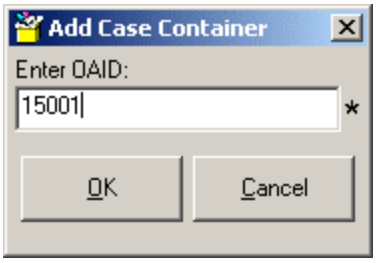
The archivist is returned to the *Systematic Case Mngmt* activity with the Systematic Case Number highlighted in the left pane and the Systematic Case properties in the right pane.

9.2 Add Systematic Case Container

Typically, Systematic Processing is performed on a series of records that consist of one or more containers. The archivist selects the Description Activity in ART to find the container ids (OAIDs) of the containers in a series and associates those container ids with a systematic case. To associate a container with a systematic case, the archivist first selects a systematic case. Once a systematic case is highlighted, the archivist selects *Add Container* from the *Edit* drop-down menu.



This causes the Add Case Container dialog box to appear.



The archivist enters the OAID of the container that is to be associated with this systematic case. By repeatedly selecting Add Container, one can add all containers associated with a series to a systematic case.

9.3 Systematic Processing of Containers

When the archivist is ready to start systematic processing of a container, they must check out the container and open it for processing. This accomplished in the same manner as for FOIA processing of a container. The only difference is that to check out or check in a container, or to undo a checkout, the archivist must be in the *Systematic Case Mngmt* activity. The operations are described in section 5 of this report.

10. Summary and Related Research

During Phase I of the Presidential Electronic Records PilOt System (PERPOS) Project, tools were prototyped that support archivists at Presidential Libraries in systematic processing of Presidential electronic records. The tools include: (1) Archival Repository Tool (ART) that supports accession, description, and creation of reference copies of opened and redacted electronic records, and (2) the Archival Processing Tool (APT) that supports arrangement, preservation and review of electronic records. The APT also supports Filtering of electronic records contained in file systems. This function involves reliably removing operating system files, office application files and software documentation files from an accessioned file system leaving just the user-created electronic records. It is comparable to the preservation activity for paper records of removing non-records from a box of paper records, e.g., books, and transferring them to the library. These tools were pilot tested by archivists at the Bush Presidential Library and functions refined to better support systematic processing [Spencer et al 2005].

During Phase II of the PERPOS project, capabilities to support FOIA processing were added to ART and the APT and those capabilities are described in this report. Those capabilities include:

- Indexing accessioned electronic records,
- Creating a FOIA case,
- Searching the indexed records for records relevant to a FOIA Request,
- Automatic estimation of the number of pages of e-records relevant to a request,

- Reviewing records for a FOIA case,
- Creating the Scope and Content Note for a FOIA case,
- Automatically creating a container for a FOIA Collection and its finding aid, and
- Managing systematic cases

There are improvements and additions that should be made to ART and the APT as regards FOIA Processing.

The Page Estimation method should be improved. The distribution of actual page sizes around the average number of bytes per page for a record of a particular file format is large. A better method would be to use a TIFF or PDF printer driver that can create multipage Tiff or PDF files from a viewer application by simply selecting the driver as the 'printer' in the Print menu. One could then access the field indicating the number of pages and associate its value with a numpages attribute in the container manifest. Furthermore, page estimation for files of various file formats is needed not just for estimating FOIA processing (review) workload, but for recording archivist productivity and for indicating the volume of records in unprocessed and systematically processed record series

In section 8 of this report, the disposition of inactive FOIA collections was discussed. It was suggested that when creating a FOIA collection, records that are related to a FOIA request that have already been processed as part of other FOIA collections, not include references to those collections where the records can be found, but include a copy of the records themselves. References in the FOIA collection and Finding Aid could still be made to systematically processed record series, or copies of those records could also be included in the FOIA collection.

This would eliminate the problem that occurs when a FOIA collection becomes inactive, of ensuring that all FOIA collections that refer to records in a collection being broken down refer to another FOIA collection or systematically processed record series where the record can now be found. Furthermore, it would be much more convenient for the researcher if more records in a FOIA collection that pertain to his request were in a single container. He would no longer get some of the relevant records and have to refer to other FOIA collections to see the remainder.

In section 7.4, it was pointed out that a Finding Aid for a FOIA Collection of paper records indicates records that are included that are not relevant to the FOIA request, but were incidentally processed. It is possible for an archivist to review records whose filename was not in boldface, and thus may or may not be relevant to the request. Currently, the Archival Repository does not keep track of or include records that were not relevant or that were incidentally processed in either the FOIA Reference Container or the Finding Aid. However, this capability can be added.

Currently, withdrawal sheets that replace closed records, and redacted records that replace original records retain their respective filenames. A filename is a type of metadata and may indicate something about the content of the file. Perhaps that

filename should not be publicly accessible, because it could indicate information about the contents that should be restricted. Hence, a capability should exist to replace the original filename with a content-free name.

When it is necessary to reindex holdings, it is currently necessary to reindex the contents of every filtered container in holdings. It would be less consuming of processor time if the containers could be incrementally indexed. That is to say, only new filtered accessions, and only containers that had records that were transferred out would need to be reindexed. Furthermore, it would be better if containers that had been filtered and containers that had other transfers of records were incrementally indexed in the background after they were checked back into holdings. This would eliminate the obstacle of an archivist discovering that they needed to reindex holdings in order to perform a FOIA search.

While ART can create reference copies and finding aids for systematically processed record series and FOIA collections, the only way that they can currently be viewed is in the archivist's Archival Repository Tool. A reference tool is needed for the Presidential Libraries that could be used by the researcher in the reference room. At a minimum, this tool would include a catalog of e-collections that had been opened for public access, and that included the capability to browse the catalog and open containers associated with systematically processed record series and containers of FOIA collections. It could also contain the capability to search for records in the Library of opened Presidential e-records.

Such an interactive catalog and access tool is easy to construct from existing components of ART and the APT. A Reference Catalog would have a structure similar to the Repository Catalog in the Description Activity of ART and the access and viewing capability correspond to the *Explore* Activity of the APT.

Archivists at the Bush Presidential Library have not yet used the tools for FOIA processing. The most recent version of the tools that support FOIA processing will be installed at the Bush Presidential Library so that archivists can pilot test them. Based on feedback from archivists assigned to pilot test the tools, refinements can be made to the tools to ensure that they effectively support FOIA processing of Presidential e-records.

To ensure that adequate audit information is being maintained for FOIA (and Systematic) Processing, there should be a systematic walkthrough of the series of actions supported by ART and the APT to determine the audit information currently collected and any additional audit information needed. This audit walkthrough should include the prototype developer and archivists.

In other research, advanced decision support functions are being developed that have the potential to substantially reduce the work that archivists at Presidential Libraries have to perform in processing Presidential e-records. For instance, a capability to automatically recognize record types such as correspondence, memoranda, schedules, and press releases, will make it possible to automatically fill in withdrawal information such as

record type, chronological date, correspondents, and subject [Harris and Underwood 2005].

The feasibility of automatically identifying Presidential Record Act restrictions and FOIA exemptions to public release of Presidential records is also being investigated [Underwood and Harris 2005]. Potential benefits of this tool include estimation of the complexity of reviewing a FOIA case and reducing the likelihood of a FOIA exemption or PRA restriction being missed during review.

The feasibility of automatically generating Scope and Content Notes for unprocessed and systematically processed record series and for FOIA collections is being investigated [Underwood 2005]. Success would provide archivists with better intellectual control over accessioned, unprocessed record series, and would reduce the effort needed to describe record series after processing.

Oracle Text with Word Query is used in ART to support FOIA search. The average precision of text-based search on large collections of documents using tools such as Oracle has been shown in the Text Retrieval Conferences to be at best .47 [Voorhees and Harman 1999]. Average precision is a good measure of the utility of a document retrieval system. Average precision combines precision, relevance ranking and overall recall. Average precision is the sum of the precision at each relevant document in the result set divided by the total number of relevant documents in the collection.

A novel approach to document retrieval is being investigated that has the potential to improve the average precision of search for records relevant to a FOIA request. The approach involves natural language Boolean queries using an index of the subjects in a collection supplemented with general conceptual knowledge and collection specific knowledge acquired from the collection itself [Iwanska and Underwood 2006]. Experiments will be conducted to determine whether this approach achieves performance significantly better that Oracle Text with word search. The significance of this research is that in responding to FOIA requests, better precision and recall in the search of e-record collections means fewer irrelevant records that need to be reviewed, so more time for archivists to process other FOIA requests. It also means faster response to requestors, more records relevant to the request and fewer that are irrelevant.

References

[Harris and Underwood 2005] B. Harris and W. E. Underwood. Induction of the Documentary Form of Record Types. PERPOS Working Paper 05-8, 2005.

[Iwanska and Underwood 2006] L. Iwanska and W. E. Underwood. Natural Language Boolean Queries. Working Paper 06-01, February 2006

[Spencer et al 2005] L. Spencer, S. Oriabure and W. Underwood. Launching E-Records with a PERPOS: The Presidential Electronic Records PilOt System. NAGARA Annual Meeting 2005. Richmond, Virginia. July 20-23, 2005

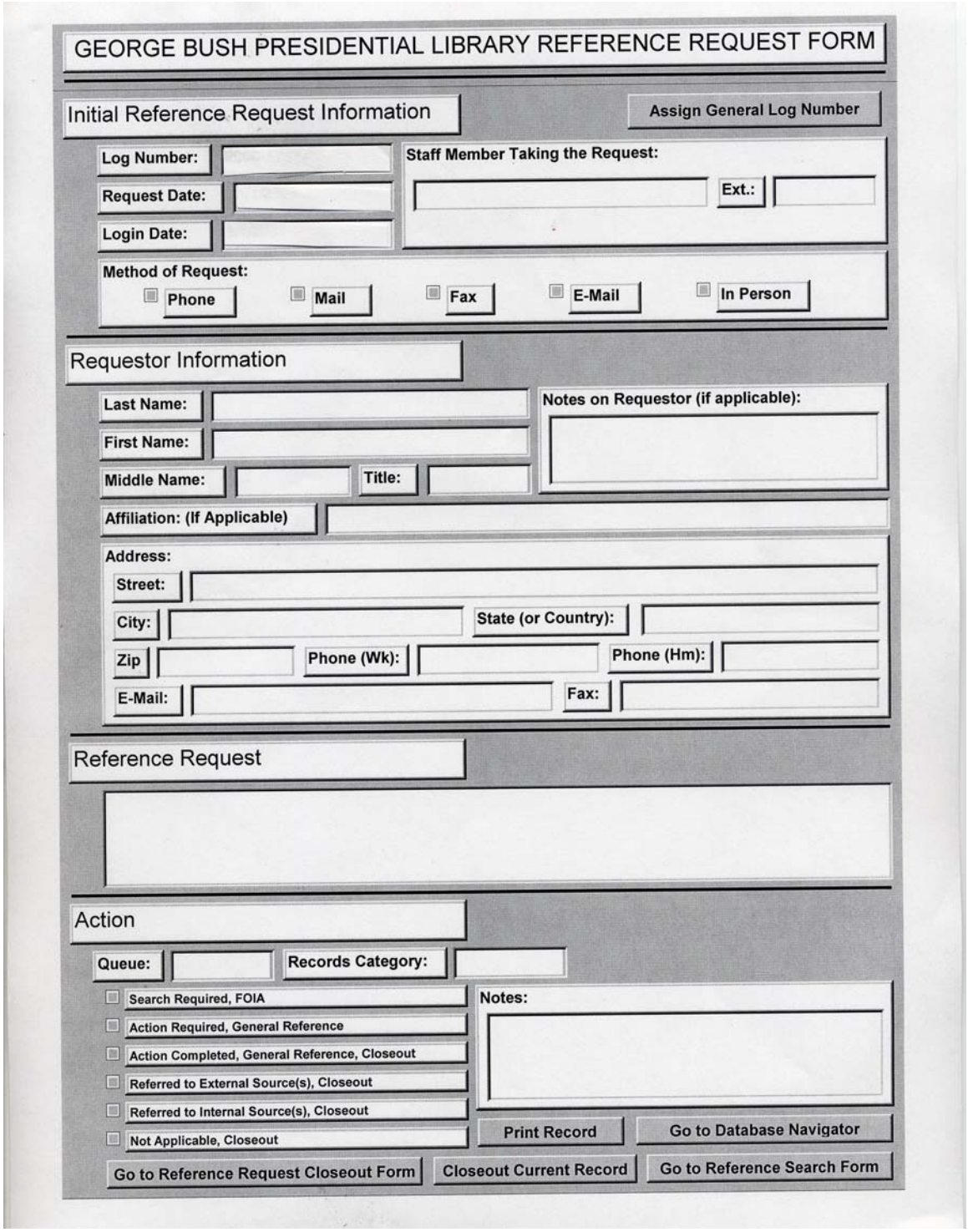
[Underwood 2006] W. E. Underwood. Automatic Description of the Content of Presidential Record Series, Working Paper 05-09, July 2006.

[Underwood and Harris 2005] W. E. Underwood and B. Harris. The Knowledge and Reasoning Required to Recognize Presidential Record Act Restrictions and Personal Record Misfiles. PERPOS Working Paper 05-03, ITTL/CSITD, Georgia Tech Research Institute, 2005.

[Underwood et al 2006] W. Underwood, M. Hayslett and S. Laib. Reference Manual for PERPOS: An Electronic Records Repository and Archival Processing System, Version 3.1. PERPOS Technical Report ITTL/CSITD 06-2, Information Technology and Telecommunications Laboratory, Georgia Tech Research Institute, February 2006.

[Voorhees and Harman 1999] Voorhees, E. M. and Harman, D. Overview of the Eighth Text Retrieval Conference (TREC-8) (1999) The Eighth Text Retrieval Conference (TREC-8), NIST Special Publication 500-246.
http://trec.nist.gov/pubs.html

Appendix A: Library Reference Request Form



GEORGE BUSH PRESIDENTIAL LIBRARY REFERENCE REQUEST FORM

Initial Reference Request Information

Assign General Log Number

Log Number:

Request Date:

Login Date:

Staff Member Taking the Request:

Ext.:

Method of Request:

Phone Mail Fax E-Mail In Person

Requestor Information

Last Name:

First Name:

Middle Name:

Title:

Notes on Requestor (if applicable):

Affiliation: (If Applicable)

Address:

Street:

City:

State (or Country):

Zip

Phone (Wk):

Phone (Hm):

E-Mail:

Fax:

Reference Request

Action

Queue:

Records Category:

Search Required, FOIA

Action Required, General Reference

Action Completed, General Reference, Closeout

Referred to External Source(s), Closeout

Referred to Internal Source(s), Closeout

Not Applicable, Closeout

Notes:

Print Record

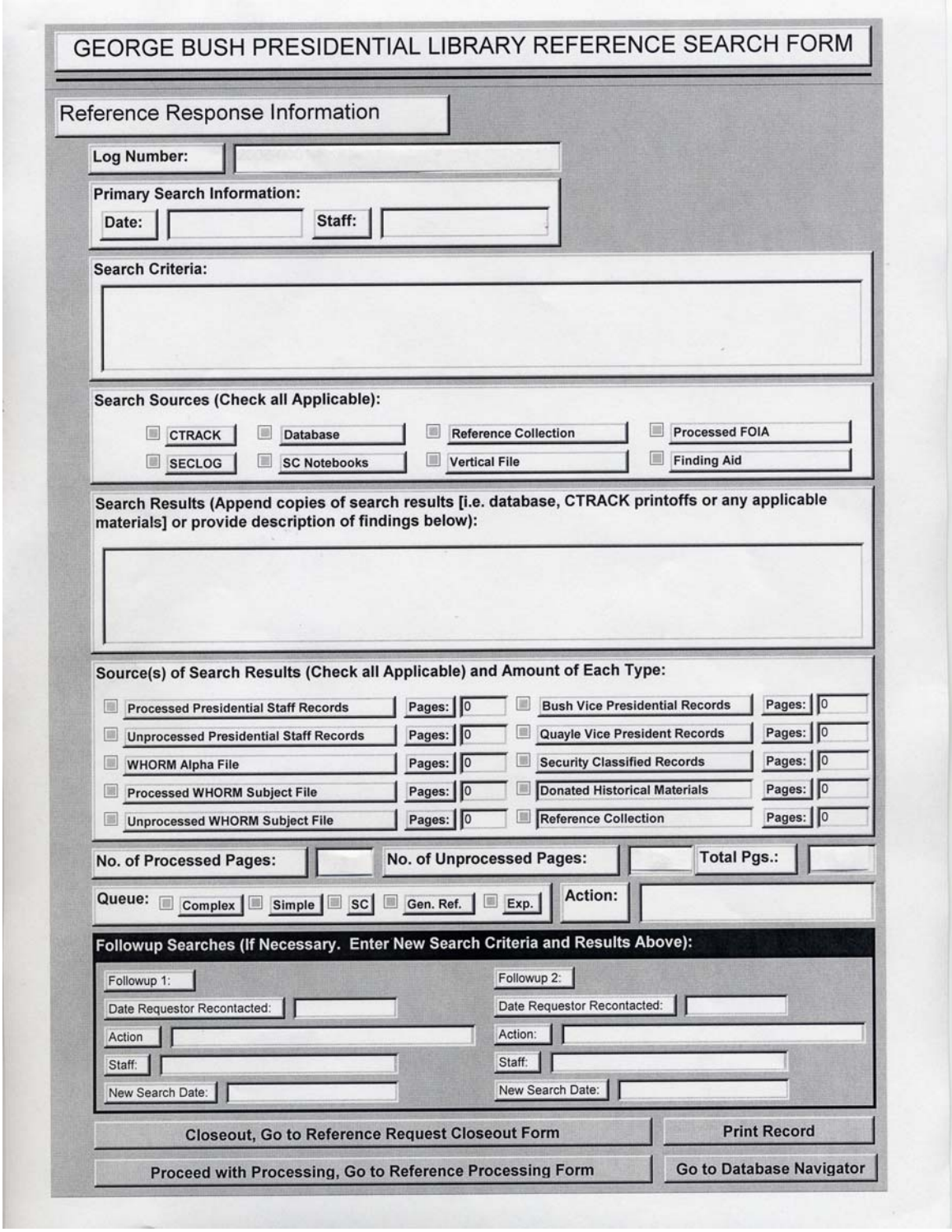
Go to Database Navigator

Go to Reference Request Closeout Form

Closeout Current Record

Go to Reference Search Form

Appendix B: Library Reference Search Form



GEORGE BUSH PRESIDENTIAL LIBRARY REFERENCE SEARCH FORM

Reference Response Information

Log Number:

Primary Search Information:

Date: Staff:

Search Criteria:

Search Sources (Check all Applicable):

CTRACK | Database | Reference Collection | Processed FOIA

SECLOG | SC Notebooks | Vertical File | Finding Aid

Search Results (Append copies of search results [i.e. database, CTRACK printoffs or any applicable materials] or provide description of findings below):

Source(s) of Search Results (Check all Applicable) and Amount of Each Type:

Processed Presidential Staff Records — Pages: 0		Bush Vice Presidential Records — Pages: 0	
Unprocessed Presidential Staff Records — Pages: 0		Quayle Vice President Records — Pages: 0	
WHORM Alpha File — Pages: 0		Security Classified Records — Pages: 0	
Processed WHORM Subject File — Pages: 0		Donated Historical Materials — Pages: 0	
Unprocessed WHORM Subject File — Pages: 0		Reference Collection — Pages: 0	

No. of Processed Pages: | No. of Unprocessed Pages: | Total Pgs.:

Queue: Complex | Simple | SC | Gen. Ref. | Exp.

Action:

Followup Searches (If Necessary. Enter New Search Criteria and Results Above):

Followup 1:	Followup 2:
Date Requestor Recontacted:	Date Requestor Recontacted:
Action	Action:
Staff:	Staff:
New Search Date:	New Search Date:

Closeout, Go to Reference Request Closeout Form | Print Record

Proceed with Processing, Go to Reference Processing Form | Go to Database Navigator

Appendix C: Sample Finding Aid for a FOIA Case

1998-0091-F

SCOPE AND CONTENT NOTE

This FOIA request contains materials on the development and implementation of the Bush Administration's policy with respect to Acquired Immune Deficiency Syndrome (AIDS) and Human Immunodeficiency Virus (HIV). The FOIA collection contains correspondence from constituents concerning personal struggles with AIDS, AIDS and immigration, AIDS spending, AIDS research, and AIDS support groups. The material in this collection consists of Bush Presidential and Quayle Vice Presidential Staff and Office Files. These files include information on AIDS related events, hemophilia and AIDS, immigration reform, the development and implementation of the Bush Administration AIDS policy, news clippings about AIDS, and reports and surveys on AIDS.

Please note that a "*" indicates that a collection is entirely processed and open. A more detailed description of the materials is available in the finding aid for the office or category.

A "**" indicates files or documents that were incidentally processed as part of this FOIA request.

BUSH PRESIDENTIAL RECORDS: WHITE HOUSE STAFF AND OFFICE FILES

Advance Office

Peggy Hazelrigg Files- Chron File

Address AIDS Coalition - Thursday, March 29, 1990 [OA/ID 06023]
Briefing with AIDS Commission - Monday, December 9, 1991 (HHS) [OA/ID 06025]

John Herrick Files - Chron File

Briefing with AIDS Commission - Monday, December 9, 1991 [OA/ID 08135]

Office of Cabinet Affairs

Daniel Casse Files

HIV/Immigration [OA/ID 07122]
President's Briefing on AIDS (HHS) [OA/ID 07123]
HIV Infection Waiver [OA/ID 07133]
AIDS Commission [OA/ID 07133]
National Commission on AIDS 12/6 [OA/ID 07134]
Meeting with National AIDS Commission - Monday, December 9, 1991 [OA/ID 08298]

Stephanie Dance Files

AIDS [OA/ID 04146]

Michael P. Jackson Files

Meeting with the AIDS Commission - Thursday, March 29, 1990 [OA/ID 06380]
AIDS Meeting in Oval Office 8/13/92 [OA/ID 06384]

Paul Korfonta Files
AIDS [OA/ID 03776]

Jay Lefkowitz Files
AIDS [OA/ID 07867]

Chief of Staff
John Sununu Files
AIDS [OA/ID CF00470]

Correspondence Office
Joan Gibson Files
[Pediatric AIDS Awareness Week Proclamation] [OA/ID 06232]
[Pediatric AIDS Awareness Week 1991 Proclamation] [OA/ID 06559]

Council of Economic Advisors
Richard Schmalensee Files
Working Group on Health Policy - AIDS [OA/ID 03686]

Richard Schmalensee Files - Subject File
Health - AIDS [OA/ID 03688]

QUAYLE VICE PRESIDENTIAL RECORDS: WHITE HOUSE STAFF AND OFFICE FILES

Competitiveness Council
[Staff Person Unknown]
AIDS - INS/HHS (Aliens) Rule [OA/ID 21931]

Counsels Office
Fischer Files - Subject File
AIDS [OA/ID 21890]

Fischer Files - Legal / Subject Matter Files - Residence
CJR [Civil Justice Reform]: AIDS Liability [OA/ID 22216]

Appendix D: Changes to the Repository Database Tables

In the earlier version of the PERPOS Tools, all processing was systematic. The only tracking that was done was whether a container had been copied to Working Storage for processing. This was accomplished with the use of the Status property of the Container table. If the Status was 'Null', the container was unprocessed and resided in the Holdings area. If the Status was 'I', the container was in process and resided in the Working Storage area. If the Status was 'P', the container was fully processed and resided in the Holdings area. No record was kept of who copied it to Working Storage or who moved it back to the Holdings area.

To enable FOIA processing, more tracking information is necessary. The Container table has been modified to store values including the name of the archivist who checked out a container and for which process it was checked out. Five new tables have been added to keep track of FOIA cases and systematic processing cases. The new tables are the FOIACase, SystematicCase, CaseContainer, CaseDocument, DocumentContainer, and the PageEstimates tables.

Two additional tables have been added. They are associated with the ManifestLibrary component.

FOIACase

The FOIACase table keeps track of information about individual FOIA cases. This is the main table for FOIA processing. The following table describes the fields in the FOIACase table. The FOIACase table is displayed in CaseNo order.

FieldName	Description
ID	This field is the primary key and is used in joins
CaseNo	FOIA Case number assigned when the case is created
SubCaseNo	Subcase number when a single FOIA request is split into subcases
Date	The date the FOIA Case was created. This field is collect automatically from the system date, when a FOIA Case is created.
Archivist	Computer UserID of the Archivist assigned to the FOIA Case. Containers under this FOIA request can only be open for edit, checked out for processing, or checked in after processing by this Archivist or the Administrator. This field is collected when the FOIA Case is created in ART.
Note	Scope and Content Note

SystematicCase

The SystematicCase table keeps track of information about individual Systematic cases. This is the main table for Systematic processing. The SystematicCase table is displayed in CaseNo order. The following table describes the fields in the SystematicCase table.

FieldName	Description
ID	This field is the primary key and is used in joins
CaseNo	Systematic Case number assigned
Archivist	Computer UserID of the Archivist assigned to the Systematic Case. Containers under this Systematic Case can only be opened for edit, checked out for processing, or checked in after processing by this Archivist or the Administrator. This value is collected when the Systematic Case is created in ART.

CaseContainer

The CaseContainer table is used to tie a container to either a Systematic case or a FOIA case. In the database design, there is a one-to-one relationship between Systematic cases and containers. There is a one-to-many relationship between a container and FOIA cases. A single container can belong to a Systematic case and/or one or more FOIA cases. The following table describes the fields in the CaseContainer table.

FieldName	Description
ID	This field is used as the primary key
CaseType	Specifies the type of case to which a container belongs. It is ‘F’ for a FOIA Case and ‘S’ for a Systematic Case. This is used as a filtering field in queries that get all containers belonging to a single case, since the CaseID is used and the same CaseID can exist in the SystematicCase table and the FOIACase table.
CaseID	The ID of a case in either the FOIACase table or a case in the SystematicCase table. This field helps tie a container to a case.
ContainerID	The ID of the container in the Container table. This field ties a container to a case.

Container

The Container table has been modified to keep track of who has checked out a container along with the type and case no of the case under which it was checked out. In the past, if a container’s Status was ‘I’ for "in process," it was in some archivist’s working storage area. It was not returned to the Holdings area until processing was complete and a

Reference Copy had been made. This meant that any container with a status of ‘U’ for Unprocessed or ‘P’ for Processed was in the Holdings area.

Now, a container can be partially processed and returned to the Holdings area. A status of ‘I’ no longer means a container is in working storage. When, a container is checked out for processing it receives an ‘I’ for In process in the Status field, the UserId of the archivist in the Archivist field, an ‘s’ for a Systematic case or ‘f’ for a FOIA case in the CaseType field, and the CaseNo of the case it was checked out under in the CaseNo field. When a container is checked back into the Repository, the Archivist, CaseType, and CaseNo are set to null. The Status field value remains ‘I’ until the container has been completely processed. If a case has never been checked out, its Status is ‘U’ for Unprocessed and the container is considered to reside in the Holdings area. When a container’s InprocessStatus become ‘FAPRD’, which means the entire container has been ‘F’ for Filtered, ‘A’ for Arranged, ‘P’ for Preserved, ‘R’ for Reviewed, and ‘D’ for Described. Its Status automatically becomes ‘P’ for Processed.

The following table describes the fields in the Container table.

FieldName	**Description**
ID	This field is used as the primary key.
OAID	This is the OA/ID assigned to the container when it is accessioned
AccessionID	This is the ID of the Accession to which the container was added. This field ties the container to the Accession table
LocationID	This is the ID in the Location table that describes where this container resides. This field is used mainly for containers that are store on secondary storage devices. This field ties the container to the Location table
ContainerType	Identifies type of container
SeriesID	Ties container to the Series Table
Name	File Name or Box Number
Arrangement	Arrangement of the contents of this container
Notes	Scope and Content Notes
Status	Values: Null (Unprocessed), P (Processed), I (In process)
HasReferenceCopy	True if a Reference copy has been created for this container
InProcessStatus	Represents the parts of processing that have been completed on the whole container; values can be a combination of FAPRD, F - Filtered, A – Arranged, P – Preserved, R – Reviewed, and D – Described.
ProcessingArchivist	UserId of the Archivist who has checked out this container
ProcessingType	The type of processing for which the container has been checked out.

CaseNo	The FOIA or Systematic case under which this container has been checked out.

ContainerDocument

The ContainerDocument table is used to store information about records that are part of one or more FOIA cases. This table makes it possible in one location to update when a record has been reviewed. The records in this table are tied to a case through the CaseDocument table. The following table describes the fields in the Container table.

FieldName	Description
ContainerID	This is ID of the container of which this record is part
ContainerDocID	This is ID of the record within the container. This field is equivalent to the ItemID in the Manifest, and ties this record to the manifest of the container with the above ContainerID
GroupID	This is the same GroupID as in the Manifest and is used with the GroupID in the Manifest
Path	This is the Path of the record. It is updated from the Manifest whenever a container is checked back in
ProcessedType	This is true if this record has been reviewed. It is updated whenever the container is checked back in. This field is also used to count of how many file in a FOIA case have been processed.
FOIARef	This field is given the FOIACase CaseNo of the first FOIA case for which a FOIA Reference Copy is made that contains this record
FileType	This field is used in page estimating. It ties to the FileType field in the PageEstimates table.
FileSize	This field is also used in page estimation.
PageEstimate	This is the estimate of the number of pages in this file.

CaseDocument

The CaseDocument table is used to tie a record to a FOIA case. This table represents a ‘one to many’ relationship between a record and one or more FOIA cases. This table also represents a ‘one to many’ relationship between a case container and many records. A single record can belong to one or more FOIA cases. A single case container can contain many records. The following table describes the fields in the CaseDocument table.

FieldName	Description
ID	This field is the primary key and is used in joins
CaseContainerID	This is the ID of the CaseContainer to which this record belongs. The CaseContainer is what ties the container to a single case.
DocContainerID	This is the ID of the container to which a record belongs.

DocID	This is ID of the record within the container. This field is equivalent to the ItemID in the Manifest, and ties this record to the manifest of the container with the above DocContainerID
GroupID	This is the same GroupID as in the Manifest.

PageEstimates

The PageEstimates table is used to calculate the estimated number of pages in a file. Different Filetypes have different file header length and different average bytes per page. This table is used to calculate the estimated number of pages a file contains given its file type.

FieldName	**Description**
FileType	A specified file type.
HeaderLength	The length of the file header for this file type.
BytesPerPage	Average no of bytes per page for this file type.

APPENDIX E: Changes to ART and the APT to Enable FOIA Processing

New capabilities, database tables, fields, and components are required, to support FOIA processing. One of the basic capabilities needed is searching the holdings for records responsive to a FOIA request. This means that the capability to index and search the contents of the files in each container in holdings is necessary. In addition, in FOIA processing, individual records rather than entire containers are processed. This means there is more information that needs to be collected and shared between the Archival Repository Tool and the Archival Processing Tool. This section describes the changes necessary to support FOIA processing.

To enable FOIA searching, the ability to index the contents of all the accessioned and filtered containers residing in Holdings has been added to the Archival Repository Tool. Archive files, e.g., zip, tar, that have not been expanded, password-protected files, image files, and audio files are indexed. Once indexing is completed, searching can occur. To add the capability to search, a new application was created. This application is called FOIASearch.

The FOIASearch application is usually called from inside the Archival Repository Tool, though it can be used as a stand-alone application. The indexing and search components call Oracle Text functions. These functions provide Boolean queries with ranking of results. The result of the search is displayed in the form of a table. One of the fields in the table is the relevance ranking of the resulting records in relation to the input query. The FOIASearch application provides the archivist the ability to select and view records that are part of the result set. The records are displayed in html format with the search terms highlighted. The archivist can limit the result to records with a ranking that is greater that one that the archivist has selected. When this application is called from inside the Archival Repository Tool, it returns both the input query and the resulting table to the Archival Repository Tool for further processing. When run as a stand-alone application, the result table is returned as a comma delimited text file with the first row containing the column headings.

The FOIASearch application displays only the Rank, OAID, and the FilePath of the files it finds relevant to the query. There are additional fields returned in the result that are used internally by other components. They are the ID and the DocID fields. The ID field uniquely identifies the record within ORACLE. It is used by the FOIASearch application to allow the archivist to view a selected record. The DocID field uniquely identifies the record within its container. It is equivalent to the ItemID field in the manifest and is used by the Archival Repository Tool to identify records that may have undergone or will undergo some form of processing that can change the path.

The query and the results are saved in the Microsoft access tables named FOIACase, CaseContainer, CaseDocument and ContainerDocument. The query is saved along with its FOIA case in the FOIACase table. A new CaseContainer records is created for each

container that contains one or more relevant records. A new CaseDocument record is created for each relevant record returned in the result. One record may occur several times in the CaseDocument table if it belongs to several different FOIA cases. A new ContainerDocument record is created for any record returned in the result set that does not already exist in the ContainerDocument table. A record can only occur one time in this table. The ProcessedType field in the ContainerDocument table provides a guide to determine what records have been processed either systematically or in a previous FOIA case and what records need to be processed. This field is used to calculate the 'Processed Files' and 'Unprocessed Files' that are displayed in the right pane when a FOIA case is highlighted. The FOIARef field is also in the ContainerDocument table and is updated the first time a record is part of a FOIA Reference container.

Another new component has been added called the ManifestLibrary. This component was created to enable the Archival Repository Tool to have access to the manifest information without having to load the entire container into the Archival Processing Tool. Previously, to gain access to information in a container's manifest, it was necessary to load the container by making a call to the APT. It is still necessary to call the Archival Processing Tool to view the contents of a container, but it is no longer necessary to load the container into the Archival Processing Tool to simply access or change manifest information such as whether the whole container has been processed or to check if an individual record has been reviewed.

This new component automatically adds three new tables to the ART database when called from ART and to the temporary database of container information when called from inside the APT. These tables are loaded with the metadata information stored in the container's manifest and the offset of the files in the container. This component may also be useful in the future when it is necessary to make batch changes to containers containing older version of a manifest or there are new File Types that are identified. The FileType field in the manifest could be blanked out for all records so that the next time the container was opened in the Archival Processing Tool, the File Types for all the records would be re-identified using the new FileType definitions.

The manifest information that was being stored in the FileItems table of the temporary database that is created each time the Archival Processing Tool loads a container has been removed. That information is now stored in a Manifest table and an Action table that the Manifest object adds to the temporary database. This was done so that any changes made to the manifest in the future only occur in one place. This change also allows both the Manifest and the FileItems objects to create stored procedures that reference the Manifest and Action tables. Manifest information is temporarily added to the Archival Repository Tool database when it is necessary to update the Manifest, the ContainerDocument, or the CaseDocument tables. The manifest information is added in the form of the Manifest and Action tables. There is one more new table called the TarItem table. It contains the HeaderOffset which is the location of a file or directories TarHeader, the TarLength which is the file size including the length of the TarHeader and any additional padding (this value is always a multiple of 512), the FileSize which is the size of the file alone, and the Path of each file. After the manifest has been read, this

temporary table is used to update the Manifest table with the TarHeader offset, the TarLength, and the FileSize. The TarHeader offset and the TarLength is used when creating a FOIA Reference container. The FileSize and the FileType fields along with a new table that has been added to ART are used to estimate the number of pages in a FOIA Case result set.

Additional changes have been made to the manifest and to the information that it contains. The manifest version has changed to four. New properties have been added for both records and directories. Both records and directories now have static ItemID, TarOffset, TarLength and a temporary CaseDoc property. Records also have a static GroupID property. The ItemID property allows the Archival Repository Tool, Archival Processing Tool, and FOIASearch to reference the same record even if the path has been changed in the Archival Processing Tool during processing. These new properties aid in the communication between the different applications.

The GroupID was added to tie both records converted to a different file format and redacted records to the original record. This property only shows up on the manifest if the ItemID and the GroupID are different. This new property allows the Archival Processing Tool to correctly handle review when there is an original and a converted and/or a redacted copy of a record. The GroupID is also used when containers and records are first added to a FOIA Case after a FOIA Search. In the case that some records were converted or redacted before the FOIA Index was created, the FOIA Search may return the document ID of the converted or redacted record but not the original document's ID. Because of these possibilities, the group of records with the same GroupID as the record that the FOIA Search returned is added to the FOIA Case. The GroupID is also used after a container has been checked back in after processing, whether the processing was Systematic or FOIA.

The CaseDoc property only shows up in the manifest of the working copy of a FOIA Case container. This property was added to allow the Archival Repository Tool to mark case records and the directories that contain them before making the working copy of a FOIA case container. This new property along with the GroupID property is also used to update all the FOIA Cases that contain a record that has been processed. The CaseDoc property is removed after all updates to the Archival Repository Tool's database have been made and before the container is returned to Holdings. This allows any converted or redacted records to be added to the FOIA Cases containing the original record. This update is done whenever a container is returned to Holdings.

Six additional tables have been added to the Archival Repository Tool database and the Series and Container tables have been modified. The Container table was modified to allow the storage of the date it was last indexed and when files where last transferred out of it. This information is used to inform the archivist when it is necessary to recreate the index of holdings before performing a FOIA Search. The table has also been modified to store temporary information about the archivist who has checked it out and the processing case under which it was checked out. The six additional tables are the SystematicCase, FOIACase, CaseContainer, CaseDocument, ContainerDocument, and

the PageEstimates tables. The SystematicCase table was added because, although all processing is no longer systematic, some of it is. This table keeps track of the name of the archivist assigned to a systematic case. This table along with the CaseContainer table keeps track of which containers were assigned to any given systematic case. The FOIACase table was added to keep track of who is assigned to a FOIA case, any Content and Scope Notes, when the case was started, the search criteria, the number of unprocessed pages, and the number of processed pages. This table along with the CaseContainer keeps track of which containers are involved in any given FOIA case. In Systematic Processing, the CaseContainer table ties multiple containers to a single systematic case. The containers are manually added to the systematic case. In FOIA Processing, the CaseContainer table is used to tie multiple containers to multiple FOIA Cases. In some cases, it is possible for a single container to be tied to a single systematic case and several FOIA cases at the same time. The containers are added to the FOIA case automatically as a result of a FOIA search. The CaseDocument table along with the CaseContainer table is used to tie a single record to one or more FOIA cases.

The ContainerDocument table allows a case record that is contained in several FOIA Cases to be changed in one place once a container is returned to Holdings after processing or after it becomes part of a FOIA Reference container. This table is used to keep track of the status of a record that belongs to one or more FOIA cases. The field used for status is called ProcessedType, since it specifies the type of processing that was done to the record. It can be an ‘f’ for records that have been FOIA processed, ‘s’ for records that have been Systematically processed, or be left blank for unprocessed records. The ContainerDocument table also contains a FOIARef field. This field only receives a value when a record is placed in a FOIA Reference container. This allows a copy of the record to be kept with the FOIA Case that will open first. If a FOIA Case includes a record that has a FOIA Ref assigned to it, it is referenced in the finding aid and is replaced with a FOIA Marker in the FOIA Reference container of all subsequent FOIA Cases. It may not be necessary to use the FOIA Marker in place of the record, since the actual record is not placed in the FOIA Reference container as it is in the paper world. Three additional fields in the ContainerDocument table are the FileType, the FileSize, and the PageEstimate fields. A copy of the open or redacted record or a withdrawal sheet is placed in a FOIA Reference container. The original of the open, closed, and redacted record are left in the Master container and copies of the records or withdrawal sheets are placed in the Reference copy of the systematically processed containers. Since copies exist in the first FOIA Reference and Systematic Reference container to be opened, there appears to be no reason why copies cannot be in every FOIA Reference container.

The PageEstimates table contains FileType, HeaderLength, and BytesPerPage. The number of pages in an electronic record is estimated by taking the FileSize, subtracting the HeaderLength associated with its FileType, and then dividing the result by the BytesPerPage associated with its FileType. The estimation method is still a research issue so the values is this table may change and there will be additions to the table. So that ART can use up-to-date information, each time "Estimate Pages" is selected, it checks for a file called “PageEstimates.txt” in its directory. This file is a comma-delimited text file. If the text file is newer than the table, the table will be automatically updated from the

text file and the PageEstimate field in the ContainerDocuments table will be updated with new calculations.

This table and these fields were added to enable ART to allow the archivist assigned to a FOIA case to select an option on the Tools menu that will calculate the estimated number of processed and unprocessed pages. It will then update the FOIA case with these values. Because these are only estimates, the archivist can chose to change these values to ones that appear more appropriate. Because files can be added to a FOIA case during processing, such as a converted copy or a redacted copy, the only files that will be used in the calculation will be the originals. If a file is added to the FOIA using the APT action menu where an archivist can add a file that may not have been indexed and therefore not found during the FOIA search, this file will become part of a future page estimate calculations.